(51) International Patent Classification 5 : C07H 15/12, C12P 21/00, 23/00 C12P 5/00, C12N 15/52, 15/70 C12N 15/74, 15/81	**A1**	**(11) International Publication Number:** WO 91/13078 **(43) International Publication Date:** 5 September 1991 (05.09.91)

(71) Applicant: AMOCO CORPORATION [US/US]; Mail Code 1907, Patents and Licensing Department, P.O. Box 87703, Chicago, IL 60680-0703 (US).

(72) Inventors: AUSICH, Rodney, Lee ; 158 Forest Street, Glen Ellyn, IL 60137 (US). BRINKHAUS, Friedhelm, Luetke ; 5885 Forest View Road #611, Lisle, IL 60532 (US). MUKHARJI, Indrani ; 2300 Central Park Avenue, Evanston, IL 60201 (US). PROFFITT, John, Houston ; 1008 N. Hayes Avenue, Oak Park, IL 60302 (US). YARGER, James, Gregory ; 221 Chasse Circle, St. Charles, IL 60174 (US). YEN, Huei-Che, Bill ; 1311 Jane Avenue, Naperville, IL 60540 (US).

(54) Title: BIOSYNTHESIS OF CAROTENOIDS IN GENETICALLY ENGINEERED HOSTS

(57) Abstract

DNA segments encoding the *Erwinia herbicola* enzymes geranylgeranyl pyrophosphate (GGPP) synthase, phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-carotene hydroxylase, and zeaxanthin glycosylase and DNA variants thereof encoding an enzyme having substantially the same biologically activity, vectors containing those DNA segments, host cells containing the vectors and methods for producing those enzymes, a method for protecting plants from the herbicide norflurazon, as well as methods for producing GGPP and the carotenoids phytoene, lycopene, β-carotene, zeaxanthin and zeaxanthin diglucoside by recombinant DNA technology in tranformed host organisms are disclosed.

Carotenoid Biosynthesis Scheme

Isopentenyl Pyrophosphate + Dimethylallyl Pyrophosphate
(IPP)   (DMAPP)

Geranyl Pyrophosphate (GPP)

IPP

Farnesyl Pyrophosphate (FPP)

IPP   *GGPP Synthase*

Geranylgeranyl Pyrophosphate (GGPP)

GGPP   *Phytoene Synthase*

Phytoene

                                 Phytoene Dehydrogenase

Lycopene

                                 Lycopene Cyclase

Beta-Carotene

                                 Beta-Carotene Hydroxylase

Zeaxanthin

                                 Zeaxanthin Glycosylase

Zeaxanthin Diglucoside

-1-

BIOSYNTHESIS OF CAROTENOIDS
IN GENETICALLY ENGINEERED HOSTS

Description

5   **Technical Field**

The present invention relates to carotenoid
biosynthesis. More specifically, this invention
relates to the isolation, characterization and
expression of the six Erwinia herbicola genes encoding
10   the enzymes geranylgeranyl pyrophosphate (GGPP)
synthase, phytoene synthase, phytoene dehydrogenase-4H,
lycopene cyclase, beta-carotene hydroxylase and
zeaxanthin glycosylase that catalyze the formation of
geranylgeranyl pyrophosphate and the carotenoids
15   phytoene, lycopene, β-carotene, zeaxanthin and
zeaxanthin diglucoside, respectively, each formed
product (through zeaxanthin) being an immediate
precursor for the next-named product. The invention
also relates to methods for expression of these Erwinia
20   herbicola enzyme genes in prokaryote hosts such as
Escherichia coli and Agrobacterium tumefaciens, in
eukaryote hosts such as yeasts like Saccharomyces
cerevisiae and higher plants such as alfalfa and
tobacco, as well as to methods for preparation of GGPP
25   and those carotenoids.

Background Art

Carotenoids are 40-carbon (C_{40}) terpenoids
consisting generally of eight isoprene (C_5) units
30   joined together. Linking of the units is reversed at
the center of the molecule. Trivial names and
abbreviations will be used throughout this disclosure,
with IUPAC-recommended semisystematic names given in
parentheses after first mention of each name.

Carotenoids are pigments with a variety of applications.

Phytoene (7,8,11,12,7',8',11',12'-ψ octahydro-ψ, ψ-carotene) is the first carotenoid in the carotenoid biosynthesis pathway and is produced by the dimerization of a 20-carbon atom precursor, geranylgeranyl pyrophosphate (GGPP). Phytoene has useful applications in treating skin disorders (U.S. Patent No. 4,642,318) and is itself a precursor for colored carotenoids. Aside from certain mutant organisms, such as Phycomyces blakesleeanus carB, no current methods are available for producing phytoene via any biological process.

In some organisms, the red carotenoid lycopene (ψ,ψ-carotene) is the next carotenoid produced of the phytoene in the pathway. For example, lycopene imparts the characteristic red color to ripe tomatoes.

Lycopene has utility as a food colorant. It is also an intermediate in the biosynthesis of other carotenoids in some bacteria, fungi and green plants.

Lycopene is prepared biosynthetically from phytoene through four sequential dehydrogenation reactions by the removal of eight atoms of hydrogen. The enzymes that remove hydrogen from phytoene are phytoene dehydrogenases. One or more phytoene dehydrogenases can be used to convert phytoene to lycopene and dehydrogenated derivatives of phytoene intermediate to lycopene are also known. For example, some strains of Rhodobacter sphaeroides contain a phytoene dehydrogenase that removes six atoms of hydrogen from phytoene to produce neurosporene.

Of interest herein is a single dehydrogenase that converts phytoene into lycopene. That enzyme removes four moles of hydrogen from each mole of phytoene, and is therefore referred to hereinafter as

phytoene dehydrogenase-4H. The Rhodobacter phytoene
dehydrogenase that removes three moles of hydrogen from
each mole of phytoene will be hereinafter referred to
as phytoene dehydrogenase-3H so that the distinctions
5   between the two enzymes discussed herein can be readily
maintained.

        Lycopene is an intermediate in the
biosynthesis of carotenoids in some bacteria, fungi,
and all green plants. Carotenoid-specific genes that
10   can be used for synthesis of lycopene from the
ubiquitous precursor farnesyl pyrophosphate include
those for the enzymes GGPP synthase, phytoene synthase,
and phytoene dehydrogenase-4H.

        Beta-carotene is the third carotenoid produced
15   in the Erwinia herbicola carotenoid biosynthesis
pathway. It is also synthesized by a number of
bacteria, fungi, and most green plants.

        Beta-carotene has utility as a colorant for
margarine and butter, as a source for vitamin A
20   production, and has recently been implicated as having
preventative effects against certain kinds of cancers.

        For example, prospective and retrospective
epidemiologic studies have consistently shown that low
levels of serum or plasma beta-carotene are associated
25   with the subsequent development of lung cancer.
Because retinol is not similarly related to lung cancer
risk, beta-carotene appears to have a protective effect
without its conversion to vitamin A. Ziegler, Amer.
Instit. Nutr., publication 022/3166/89, 116 (1989).

30   Beta-carotene is produced by the cyclization
of unsaturated carotenoids in a procedure not yet well
understood. Bramley et al, In Current Topics in
Cellular Regulation 29:291,297 (1988). Because only
mutants that accumulate lycopene but not gamma-carotene
35   (another potential precursor) have been found, it is

believed that in both plants and microorganisms a
single cyclase is responsible for conversion of
lycopene to beta-carotene. Generally, the enzymes
involved in this cyclization have been found as
5   integral membrane proteins.

Current methods for commercial production of
beta-carotene include isolation from carrots, chemical
synthesis [Isler et al., United States Patent 2,917,539
(1959)] and microbial production by hoanephora trispora
10   [Zajic, United States Patents 2,959,521 (1960) and
3,128,236 (1964)].

Zeaxanthin (β,β-carotene-3,3'-diol) is the
fourth carotenoid produced in the Erwinia herbicola
carotenoid biosynthesis pathway. Zeaxanthin, a yellow
15   pigment, is currently used as a colorant in the poultry
industry.

Chemical synthesis methods for zeaxanthin
production exist, but these are inefficient and not
commercially competitive with the existing biomass
20   sources. Presently, the commercial sources for
zeaxanthin are corn grain, corn gluten meal and
marigold petals. The level of zeaxanthin in corn
kernels averages about 0.001 percent (dry weight) and
the level in corn gluten meal averages about 0.01
25   percent (dry weight). All these sources are
characterized by low and inconsistent production
levels.

Zeaxanthin diglucoside, the fifth carotenoid
produced in the Erwinia herbicola carotenoid
30   biosynthesis pathway, is also useful as a food colorant
and has a yellow color similar to that of zeaxanthin.

Carotenoids are synthesized in a variety of
bacteria, fungi, algae, and higher plants. At the
present time only a few plants are widely used for
35   commercial carotenoid production. However, the

-5-

productivity of carotenoid synthesis in these plants is
relatively low and the resulting carotenoids are
expensively produced.

One way to increase the productive capacity of
biosynthesis would be to apply recombinant DNA
technology. Thus, it would be desireable to produce
carotenoids generally and zeaxanthin and its
diglucoside specifically by recombinant DNA technology.
This permits control over quality, quantity and
selection of the most suitable and efficient producer
organisms. The latter is especially important for
commercial production economics and therefore
availability to consumers. For example, yeast, such as
S. cerevisiae in large fermentors and higher plants,
such as alfalfa or tobacco, can be mobilized for
carotenoid production as described hereinafter.

An organism capable of carotenoid synthesis
and a potential source of genes for such an endeavor is
Erwinia herbicola, which is believed to carry putative
genes for carotenoid production on a plasmid (Thiry, J.
Gen. Microbiol., 130:1623 (1984)) or chromosomally
(Perry et al., J. Bacteriol., 168:607 (1986)). Erwinia
herbicola is a genus of Gram-negative bacteria of the
ENTEROBACTERIACEAE family, which are facultative
anaerobes. Indeed, recently published European patent
application 0 393 690 A1 (published April 20, 1990;
sometimes referred to herein as "EP 0 393 690") reports
use of DNA from Erwinia uredovora 20D3 (ATCC 19321) for
preparing carotenoid molecules.

As is discussed in detail hereinafter, the
present invention utilizes DNA from Erwinia herbicola
EHO-10 (AT 39368) for preparation of carotenoid
molecules and the enzymes used in their synthesis.
Erwinia herbicola EHO-10 used herein is also referred
to as Escherichia vulneris.

-6-

The genus is commonly divided into three
groups. Of the three, the Herbicola group includes
species (e.g. E. herbicola) which typically form yellow
pigments that have now been found to be carotenoids.

5   These bacteria exist as saprotrophs on plant
surfaces and as secondary organisms in lesions caused
by many plant pathogens. They can also be found in
soil, water and as opportunistic pathogens in animals,
including man.

10   A precise organismic function has yet to be
ascribed to the pigment(s) produced by Erwinia
herbicola. Perry et al., J. Bacteriol., 168:607
(1986), showed that the genes coding for the production
of a then unknown yellow pigment lie within an

15   approximately 13-kilobase (kb) sequence coding for at
least seven polypeptides, and that the expression of
the yellow pigment is cyclic AMP mediated. Tuveson, J.
Bacteriol., 170:4675 (1988), demonstrated that these
genes, cloned from Erwinia herbicola and expressed in

20   an E. coli strain, offered the host some protection
against inactivation by near-UV light and specific
phototoxic molecules.

          E. coli and S. cerevisiae are commonly used
for expressing foreign genes, but to optimize yields

25   and minimize technical maintenance procedures, it would
be preferable to utilize a higher plant species.

Brief Summary of the Invention

          One aspect contemplated by this invention is

30   an isolated DNA segment comprising a nucleotide
sequence that defines a structural gene or variant DNA
thereof capable of expressing each of the Erwinia
herbicola genes for GGPP synthase (including a DNA
analog), phytoene synthase, phytoene dehydrogenase-4H,

-7-

lycopene cyclase, beta-carotene hydroxylase and
zeaxanthin glycosylase in biologically active form.

Another aspect of this invention is a
recombinant DNA molecule comprising a vector

5   operatively linked to an above exogenous DNA segment
isolated from Erwinia herbicola or a variant DNA. This
exogenous DNA segment defines a structural gene capable
of expressing any one of the above Erwinia herbicola
enzymes. Also included is a promoter suitable for

10   driving the expression of the enzyme in a compatible
host organism. Exemplary, particularly preferred
vectors include plasmids pARC417BH, pARC489B, pARC489D,
pARC285, pARC140N, pARC145G, pARC496A, pARC146D,
pATC228, pATC1616, pARC1509, pARC1510, pARC1520,

15   pARC404BH, pARC406BH, pARC145H and pARC2019.

A further aspect of this invention is a method
for preparing each of the above-mentioned E. herbicola
enzymes. This method comprises initiating a culture,
in a nutrient medium, of prokaryotic or eukaryotic host

20   cells transformed with a recombinant DNA molecule
containing an expression vector compatible with the
cells. This vector is operatively linked to an
isolated exogenous Erwinia herbicola DNA segment or
variant DNA that defines the structural gene for an

25   above-mentioned particular enzyme. The culture is
maintained for a time period sufficient for the cells
to express the enzyme.

Still another aspect contemplated by this
invention is a method for producing GGPP, phytoene,

30   lycopene, β-carotene, zeaxanthin and/or zeaxanthin
diglucoside that comprises initiating a culture in a
nutrient medium of prokaryotic or eukaryotic host cells
that provides the immediate precursor to the desired
product, those prokaryotic or eukaryotic host cells

35   being transformed with one or more recombinant DNA

molecule(s) described herein that include a structural
gene that can express an enzyme that converts the
precursor to the product desired. The culture is
maintained for a time period sufficient for the host
5   cells to express the enzyme and for the expressed
enzyme to convert the provided precursor into product.
A recombinant DNA molecule contains an expression
vector compatible with the host cells operatively
linked to one or more exogenous Erwinia herbicola DNA
10   or variant DNA segments comprising (i) a nucleotide
base sequence corresponding to a sequence defining a
structural gene for geranylgeranyl pyrophosphate
synthase, (ii) a nucleotide base sequence corresponding
to a sequence defining a structural gene for phytoene
15   synthase, (iii) a nucleotide base sequence
corresponding to a sequence defining a structural gene
for phytoene dehydrogenase-4H, (iv) a nucleotide base
sequence corresponding to a sequence defining a
structural gene for lycopene cyclase, (v) a nucleotide
20   base sequence corresponding to a sequence defining a
structural gene for beta-carotene hydroxylase, and/or
(vi) a nucleotide base sequence corresponding to a
sequence defining a structural gene for zeaxanthin
glycosylase as previously described. The culture is
25   maintained for a time period sufficient for the cells
to express the enzyme products of the desired
structural genes (i), (ii), (iii), (iv), (v) and/or
(vi), and form a product that is desired.
          Yet another aspect of this invention is a
30   method of protecting a higher plant from the herbicide
norflurazon. Here, a higher plant to be protected is
transformed with a recombinant DNA molecule that
encodes a structural gene for the Erwinia herbicola
enzyme phytoene dehydrogenase-4H or a DNA variant
35   thereof that encodes an enzyme exhibiting substantially

the same biological activity. The transformed plant is
maintained for a time period sufficient for phytoene
dehydrogenase-4H to be expressed, and transformed plant
is treated with a herbicidal amount of norflurazon.

In particularly preferred practice, all of the
recombinant DNA utilized in this invention is from
Erwinia herbicola. Another preferred embodiment of
this invention is a recombinant DNA molecule as
described above, wherein the promoter is Rec 7 for
E. coli, PGK, GAL 10 and GAL 1 for yeasts such as
S. cerevisiae and CaMV 35S for higher plants.

Other preferred embodiments contemplate the
methods of preparation described above, wherein the
host transformed is either a prokaryote, such as
E. coli, a eukaryote, for example yeast such as
S. cerevisiae, or a higher plant, such as alfalfa or
tobacco. Because of the utility of GGPP and the
carotenoids as chemical precursors and as effective and
apparently harmless food colorants, the ability to
produce these materials in commercially advantageous
amounts from transgenic biological sources with the aid
of recombinant DNA technology is a major benefit
flowing from this invention. To realize these
benefits, the above aspects and embodiments are
contemplated by this invention. Still further
embodiments and advantages of the invention will become
apparent to those skilled in the art upon reading the
entire disclosure contained herein.

Brief Description of Drawings

Figure 1 is a flow diagram of the carotenoid
biosynthesis pathway utilizing the Erwinia herbicola
gene complement located in the plasmid pARC376.

Figure 2 in three sheets as Figure 2-1, Figure
2-2, and Figure 2-3 illustrates the nucleotide base

sequences of certain preferred DNA segments of the
structural gene for geranylgeranyl pyrophosphate (GGPP)
synthase (SEQ ID NO: 1). The base sequences are shown
conventionally from left to right and in the direction
5   of 5' terminus to 3' terminus, using the single letter
nucleotide base code.

The reading frame of the 5' end of the
structural gene illustrated herein is indicated by
placement of the deduced, amino acid residue sequence
10   (SEQ ID NO: 2) of the protein for which it codes below
the nucleotide sequence, such that the triple letter
code for each amino acid residue is located directly
below the three-base codon for each amino acid residue.
Numerals to the right of the DNA sequence indicate
15 -   nucleotide base positions within the DNA sequence
shown. All of the structural genes shown in the
figures herein are similarly illustrated, with amino
acid initiation position beginning here with the
initial methionine residue (Met) at DNA position about
20   124 as shown.

Several restriction enzyme sites of importance
are indicated above the DNA sequence. These represent
points of manipulation in engineering the gene
construct encoding the enzyme.

25   **Figure 3** shown in three sheets as Figure 3-1,
Figure 3-2 and Figure 3-3 illustrates the DNA (SEQ ID
NO: 3) and deduced amino acid residue (SEQ ID NO: 4)
sequences of more preferred, heterologous structural
genes of Erwinia herbicola GGPP synthase. Here, the
30   expressed protein begins with the **Met** residue at about
position 150 as shown and terminates within the Eco RV
site (about 1153) in the DNA construct present in
plasmid pARC489B, whereas the gene terminates at the
Bal I site (about 1002) in the DNA construct present in
35   plasmid pARC489D. The short amino-terminal sequence

MetAlaGluPhe (about 150-161) is a heterologous sequence
from plasmid pARC306A, and is substituted for the
native sequence from DNA position 124 to 150 shown in
Figure 2.

5   **Figure 4** shown in three sheets as Figure 4-1,
Figure 4-2 and Figure 4-3 illustrates the nucleotide
(SEQ ID NO: 5) and amino acid (SEQ ID NO: 6) sequences
of the structural gene for phytoene synthase.

         The Met initiation codon (about position 16 as
10   shown) corresponds to about position 6383 on pARC376 in
Figure 5. The Bam HI restriction site at about 1093 in
Figure 4 corresponds to the Bam HI site at about
position 5302 on pARC376 in Figure 5. The illustrated
Bgl II restriction site shown at about position 8 is
15   not present in the native DNA sequence and was added as
is discussed hereinafter.

         Figure 5 schematically illustrates the plasmid
pARC376 containing the full complement of enzyme genes,
represented by capital letters, required for the
20   synthesis of carotenoids from farnesyl pyrophosphate,
as indicated in the schematic of Figure 1. Note that
the direction of transcription (arrows) is uniform for
all enzyme structural genes except beta-carotene
hydroxylase (F), which is transcribed in an opposite
25   direction. Important restriction enzyme sites are also
identified. The synthesis of phytoene is catalyzed by
the enzymes GGPP synthase (A) and phytoene synthase
(E). The gene labeled D encodes the enzyme phytoene
dehydrogenase-4H. Genes labeled B, C and F encode the
30   enzymes zeaxanthin glycosylase, lycopene cyclase, and
beta-carotene hydroxylase, respectively. The overlap
of genes E and F is shown by hatching.

         Figure 6 is a schematic representation of the
plasmid pARC306A, which contains the Rec 7 promoter.
35   This plasmid also has multiple cloning sites adjacent

to the Rec 7 promoter and 5' and 3' transcription
termination loops. Approximate positions of
restriction enzyme sites are shown.

      Figure 7 illustrates schematically the plasmid
pARC135, which contains the S. cerevisiae
phosphoglyceric acid kinase (PGK) promoter operatively
linked at the Bgl II site. Various additional features
of the plasmid are also illustrated.

      Figure 8 shows a schematic representation of
the vector pSOC713, including a partial restriction
enzyme map.

      Figure 9 is a schematic representation of
plasmid pARC145B, which is a yeast/E. coli shuttle
vector for expression of introduced genes in yeast,
including a partial restriction enzyme map.

      Figure 10 is a schematic representation of the
vector pARC145G, which is basically pARC145B above that
contains the two preferred genes; i.e., GGPP synthase
and phytoene synthase, each operatively linked at their
5' ends to the divergent promoters GAL 10 and GAL 1.
Phytoene synthase also has a PGK terminator at the 3'
end.

      Figure 11 shown in four panels as Figure 11-1,
Figure 11-2, Figure 11-3 and Figure 11-4 illustrates
the DNA (SEQ ID NO: 7) and deduced amino acid residue
(SEQ ID NO: 8) sequences of the Erwinia herbicola
structural gene for phytoene dehydrogenase-4H. The Met
codon (shown at position 7) corresponds to position
7849 on plasmid pARC376 in Figure 5.

      Figure 12 is a schematic representation of the
vector pSOC925, including a partial restriction enzyme
map.

      Figure 13 is a schematic representation of
plasmid pARC146, including a partial restriction enzyme
map.

-13-

Figure 14 shows the vector pARC146D, including
a partial restriction enzyme map.

Figure 15 shown in four panels as Figures 15-
1, Figures 15-2, Figure 15-3 and Figure 15-4
illustrates the DNA (SEQ ID NO: 9) and deduced amino
acid residue (SEQ ID NO: 99) sequence of the Erwinia
herbicola structural gene for phytoene dehydrogenase-4H
present in plasmid pARC146D.

Figure 16 is a schematic representation of
plasmid pATC228, including a partial restriction enzyme
map. In this figure, A-F are schematic representations
of the following sequences: A=tac promoter, B=phytoene
dehydrogenase-4H gene, C=pMB1 ori, D=ampicillin
resistance gene, E=chloramphenicol resistance gene, and
F=R1162 ori.

Figure 17 illustrates the coding sequence and
encoded transit peptide (SEQ ID NO: 98) DNA (SEQ ID NO:
10) sequence linked to the 5' end of the phytoene
dehydrogenase-4H structural gene for transport of the
phytoene dehydrogenase-4H enzyme into tobacco
chloroplasts as well as other plant chloroplasts.
Stars over nucleotide positions 69 and 72 in this
sequence indicate G for T and G for A replacements
utilized to introduce an Nar I site.

Figure 18 is a schematic representation of the
about 15.6 kb plasmid pATC1616, including a partial
restriction enzyme map. In this figure, A-I are
schematic representations of the following sequences:
A=CaMV 35S promoter, B=transit peptide sequence,
C=phytoene dehydrogenase-4H gene, D=NOS polyadenylation
site, E=pBR322 ori, F=ori T, G=tetracycline resistance
gene, H=ori V, and I=kanamycin resistance gene.

Figure 19 shown in three panels as Figure 19-
1, Figure 19-2 and Figure 19-3 illustrates the DNA (SEQ
ID NO: 11) and a deduced amino acid residue (SEQ ID NO:

-14-

13) sequences of the <u>Erwinia</u> <u>herbicola</u> structural gene
for lycopene cyclase.

    The Met codon (shown at position 19)
corresponds to position 9002 on plasmid pARC376 in

5   Figure 5. The restriction sites Sph I and Bam HI were
introduced at the 5' and 3' ends of the gene using PCR.
The changes in the sequence for the genetically
engineered version of the gene (SEQ ID NO: 12) used for
expression in yeast are shown in bold underneath the

10   native sequence. At the 5' end of the gene, the native
initiation GTG codon has been changed to an ATG codon.
The second amino acid residue, Arg, was originally
encoded by an AGG codon that was changed to a CGG
codon, while retaining its coding for the Arg amino

15   acid residue.

    Figure 20 shown in three panels as Figure
20-1, Figure 20-2 and Figure 20-3 illustrates the DNA
sequence of the native <u>Erwinia</u> <u>herbicola</u> DNA (SEQ ID
NO: 14) segment containing the structural gene for

20   beta-carotene hydroxylase, corresponding to a DNA
segment from position 4886 to position 5861 of pARC376
shown in Figure 5. The Met initiation codon is located
at position 25 as shown, which corresponds to position
4991 in Figure 5.

25   **Figure 21** shown in three panels as Figure
21-1, Figure 21-2 and Figure 21-3 illustrates the DNA
(SEQ ID NO: 15) and deduced amino acid residue (SEQ ID
NO: 16) sequences of the engineered <u>Erwinia</u> <u>herbicola</u>
structural gene for beta-carotene hydroxylase. The Met

30   codon (shown at position 25) corresponds to position
4991 on plasmid pARC376 in Figure 5. The restriction
sites Nco I and Sma I were introduced at the 5' and 3'
ends of the gene as described in Example 21. The
changes in the sequence for the genetically engineered

35   version of the gene are shown in bold. At the 5' end

of the gene, the native second and third amino acid
residues have been changed from -Leu-Val- to -Val-Leu-.

Figur 22 is a schematic representation of
plasmid pARC300E, including a partial restriction
5 enzyme map showing restriction sites present only once.

Figure 23 is a schematic representation of
plasmid pARC300M, including a partial restriction
enzyme map showing restriction sites present only once.

Figure 24 is a schematic representation of
10 plasmid pARC300T, including a partial restriction
enzyme map showing restriction sites present only once.

Figure 25 shown in three panels as Figure 25-
1, Figure 25-2 and Figure 25-3 illustrates the DNA (SEQ
ID NO: 97) and deduced amino acid residue (SEQ ID NO:
15 17) sequences of the engineered Erwinia herbicola
structural gene for zeaxanthin glycosylase.

Detailed Description of the Invention

A. Definition of Terms

20   **Amino Acid:** All amino acid residues identified
herein are in the natural L-configuration. In keeping
with standard polypeptide nomenclature, J. Biol. Chem.,
243:3557-59, (1969), abbreviations for amino acid
residues are as shown in the following Table of
25 Correspondence:

-16-

TABLE OF CORRESPONDENCE

	SYMBOL		AMINO ACID
5	**1-Letter**	**3-Letter**	
	Y	Tyr	L-tyrosine
	G	Gly	glycine
	F	Phe	L-phenylalanine
	M	Met	L-methionine
10	A	Ala	L-alanine
	S	Ser	L-serine
	I	Ile	L-isoleucine
	L	Leu	L-leucine
	T	Thr	L-threonine
15	V	Val	L-valine
	P	Pro	L-proline
	K	Lys	L-lysine
	H	His	L-histidine
	Q	Gln	L-glutamine
20	E	Glu	L-glutamic acid
	W	Trp	L-tryptophan
	R	Arg	L-arginine
	D	Asp	L-aspartic acid
	N	Asn	L-asparagine
25	C	Cys	L-cysteine

It should be noted that all amino acid residue
sequences are represented herein by formulae whose left
30   to right orientation is in the conventional direction
of amino-terminus to carboxy-terminus.

        Expression: The combination of intracellular
processes, including transcription and translation
undergone by a structural gene to produce a
35   polypeptide.

Expr ssion v ctor: A DNA sequence that forms
control elements that regulate expression of structural
genes when operatively linked to those genes.

Operatively linked or inserted: A structural
5   gene is covalently bonded in correct reading frame to
another DNA (or RNA as appropriate) segment, such as to
an expression vector so that the structural gene is
under the control of the expression vector.

Promoter: A recognition site on a DNA
10   sequence or group of DNA sequences that provide an
expression control element for a gene and to which RNA
polymerase specifically binds and initiates RNA
synthesis (transcription) of that gene.

Recombinant DNA molecule: A hybrid DNA
15   sequence comprising at least two nucleotide sequences
not normally found together in nature.

Structural gene: A DNA sequence that is
expressed as a polypeptide, i.e., an amino acid residue
sequence.

20   **Vector:** A DNA molecule capable of replication
in a cell and/or to which another DNA segment can be
operatively linked so as to bring about replication of
the attached segment. A plasmid is an exemplary
vector.

25

B.   **Introduction**

Constituting the most widespread group of
pigments, carotenoids are present in all photosynthetic
organisms, where they are an essential part of the
30   photosynthetic apparatus.

Mevalonic acid, the first specific precursor
of all the terpenoids is formed from acetyl-CoA via
HMG-CoA (3-hydroxy-3-methylglutaryl-CoA), and is itself
converted to isopentenyl pyrophosphate (IPP), the
35   universal isoprene unit. After isomerization of IPP to

-18-

dimethylallyl pyrophosphate and a series of
condensation reactions adding IPP, geranylgeranyl
pyrophosphate (GGPP) is formed according to the scheme
in Figure 1. The formation of GGPP is the first step

5   in carotenoid biosynthesis.

In the bacterium _Erwinia herbicola_, phytoene
has now been found to be formed biosynthetically in a
two-step process as shown in Figure 1. The initial
step is the condensation of farnesyl pyrophosphate

10   (FPP) and isopentenyl pyrophosphate (IPP) to form
geranylgeranyl pyrophosphate (GGPP). This reaction is
catalyzed by the enzyme geranylgeranyl pyrophosphate
synthase (GGPP synthase). This first step is
immediately followed by a tail-to-tail dimerization of

15   GGPP, catalyzed by the enzyme phytoene synthase, to
form phytoene. This pathway thus differs from the
pathway reported in published European Application
0 393 690 wherein GGPP is said to form prephytoene
pyrophosphate (a cyclopropylene-containing molecule)

20   that thereafter forms phytoene.

Lycopene has now been found to be the second
carotenoid produced in _Erwinia herbicola_. The third
carotenoid produced by _Erwinia herbicola_ results from
the cyclization of lycopene to form beta-carotene, by

25   the enzyme lycopene cyclase. The fourth carotenoid in
the _Erwinia herbicola_ pathway is zeaxanthin that is
produced from beta-carotene. The fifth carotenoid,
zeaxanthin diglucoside, is produced from zeaxanthin.

The present invention relates to these steps

30   in the carotenoid biosynthesis pathway, the methods of
isolating the _Erwinia herbicola_ genes encoding
carotenoid biosynthesis enzymes of the pathway and to
the adaptation of this pathway by recombinant DNA
technology to achieve methods and capabilities of GGPP

35   and carotenoid production, particularly in host

-19-

organisms that do not otherwis synthesize those
materials, but in relatively small amounts or in
specialized locations.

The disclosure below provides a detailed
description of the isolation of carotenoid synthesis
genes from Erwinia herbicola, modification of these
genes by genetic engineering, and their insertion into
compatible plasmids suitable for cloning and expression
in E. coli, yeasts, fungi and higher plants. Also
disclosed are methods for preparation of the
appropriate enzymes and the methods for GGPP and
carotenoid production in these various hosts.

Plasmid constructs are exemplified for several
host systems. However, similar constructs utilizing
the genes of this invention are available for virtually
any host system as is well known in the art.

A structural gene or isolated, purified DNA
segment of this invention is often referred to as a
restriction fragment bounded by two restriction
endonuclease sites and containing a recited number of
base pairs. A structural gene of this invention is
also defined to include a sequence shown in a figure
plus variants of such genes (described hereinafter),
that hybridize non-randomly with a gene shown in the
figure under stringency conditions as described
hereinafter. Each contemplated gene includes a recited
non-randomly-hybridizable variant DNA sequence, encodes
a particular enzyme and also produces biologically
active molecules of the encoded enzymes when suitably
transfected into and expressed in an appropriate host.

Polynucleotide hybridization is a function of
sequence identity (homology), G+C content of the
sequence, buffer salt content, sequence length and
duplex melt temperature (T_m) among other variables.
See, Maniatis et al., Molecular Cloning, Cold Spring

-20-

Harbor Laboratory, Cold Spring Harbor, N.Y. (1982),
page 388.

    With similar sequence lengths, the buffer salt
concentration and temperature provide useful variables
for assessing sequence identity (homology) by
hybridization techniques. For example, where there is
at least 90 percent homology, hybridization is carried
out at 68°C in a buffer salt such as 6XSCC diluted from
20XSSC [Maniatis et al., above, at page 447]. The
buffer salt utilized for final Southern blot washes can
be used at a low concentration, e.g., 0.1XSSC and at a
relatively high temperature, e.g. 68°C, and two
sequences will form a hybrid duplex (hybridize). Use
of the above hybridization and washing conditions
together are defined as conditions of high stringency
or highly stringent conditions.

    Moderately high stringency conditions can be
utilized for hybridization where two sequences share at
least about 80 percent homology. Here, hybridization
is carried out using 6XSSC at a temperature of about
50-55°C. A final wash salt concentration of about
1-3XSSC and at a temperature of about 60-68°C are used.
These hybridization and washing conditions define
moderately high stringency conditions.

    Low stringency conditions can be utilized for
hybridization where two sequences share at least 40
percent homology. Here, hybridization carried out
using 6XSSC at a temperature of about 40-50°C, and a
final wash buffer salt concentration of about 6XSSC
used at a temperature of about 40-60°C effect non-
random hybridization. These hybridization and washing
conditions define low stringency conditions.

    An isolated DNA or RNA segment that contains a
nucleotide sequence that is at least 80 percent, and
more preferably at least 90 percent identical to a DNA

sequence for gene shown in a figure is contemplated by
this invention. Such a nucleotide sequence, when
present in a host cell as part of a plasmid or
integrated into the host genome as described herein,
that also hybridizes non-randomly under at least
moderately high stringency conditions, and encodes and
expresses a biologically active enzyme is contemplated
herein as a variant DNA of an illustrated sequence that
exhibits substantially the same biological activity.

In living organisms, the amino acid residue
sequence of a protein or polypeptide is directly
related via the genetic code to the deoxyribonucleic
acid sequence of the structural gene that codes for the
protein. Thus, a structural gene can be defined in
terms of the amino acid residue sequence; i.e., protein
or polypeptide, for which it codes.

Thus, through the well-known redundancy of the
genetic code, additional DNA and corresponding RNA
sequences can be prepared that encode the same amino
acid residue sequences, but that are sufficiently
different from a before-discussed gene sequence that
the two sequences do not hybridize at high stringency,
but do hybridize at moderately high stringency.
Furthermore, allelic variants of a structural gene can
exist in other _Erwinia herbicola_ strains that are also
useful, but form hybrid duplex molecules only at
moderately high stringency.

A DNA or RNA sequence that (1) encodes an
enzyme molecule exhibiting substantially the same
biological activity as an enzyme molecule expressed by
a DNA sequence of a figure, (2) hybridizes with a DNA
sequence of one of those figures at least at moderately
high stringency and (3) shares at least 80 percent, and
more preferably at least 90 percent, identity with a
DNA sequence of those figures is defined as a variant

-22-

DNA sequence. Thus, a DNA variant or variant DNA
sequence is defined as including an RNA sequence.

In comparing DNA sequences of E. herbicola and
E. uredovora, the published European Application 0 393

5   690 reported no hybridization of a DNA probe from E.
uredovora with DNA from E. herbicola using highly
stringent hybridization conditions. Present studies
indicate a range of sequence identities of about 55 to
about 70 percent between the sequences of that

10   published European application and the sequences
disclosed herein, with there being about a 65 percent
identity between the two genes for beta-carotene
hydroxylase. In view of the 45 to 30 percent of
mismatched base pairs, and the reported non-

15   hybridization at high stringency of the herbicola and
uredovora DNAs, the reported E. uredovora DNA sequences
and the E. herbicola DNAs discussed herein are
distinguishable, and genes that produce similar enzymes
are not variants.

20   Variant DNA molecules that encode and can
express a desired GGPP or carotenoid enzyme can be
obtained from other organisms using hybridization and
functionality selection criteria discussed herein. For
example, a microorganism, fungus, alga, or higher plant

25   that is known or can be shown to produce a carotenoid
is utilized as a DNA source. The total DNA of the
selected organism is obtained and a genomic library is
constructed in a λ phage such as λgt11 using the
protocols discussed in Maniatis et al., Molecular

30   Cloning, Cold Spring Harbor Laboratory, Cold Spring
Harbor, N.Y. (1982) at pages 270-294.

The phage library is then screened under
standard protocols using a radiolabeled, nick-
translated DNA probe having a sequence of the Erwinia

35   herbicola DNA of the figures, and the before-discussed

-23-

moderate or high stringency hybridization conditions.
Once the hybridization studies locate the appropriate
structural gene, that structural gene DNA segment can
be obtained, sequenced, engineered for expression in an
5   appropriate recombinant molecule and shown to produce a
biologically active enzyme as is discussed elsewhere
herein and other techniques and protocols that are well
known to workers skilled in molecular biology.

        That a DNA sequence variant encodes a
10   "biologically active" enzyme or an enzyme that has
"substantially the same biological activity" is
determined by whether the variant DNA sequence produces
an enzyme as discussed herein. Thus, a DNA variant
sequence that expresses GGPP synthase or a carotenoid
15   biosynthesis enzyme that converts a provided precursor
substrate molecule into a desired product is defined as
biologically active. Expression of a biologically
active enzyme from a variant DNA sequence can be
assayed by the production of the desired product.

20   A DNA segment of the invention thus includes a
DNA sequence that encodes Erwinia herbicola GGPP
synthase, phytoene synthase, phytoene dehydrogenase-4H,
lycopene cyclase beta-carotene hydroxylase or
zeaxanthin glycosylase of a figure, or a variant DNA.
25   In a preferred embodiment, that DNA segment includes a
DNA sequence that encodes the enzyme in a DNA segment
separate from any other carotenoid-forming enzyme
encoding sequences. More preferably, a DNA segment
contains the Erwinia herbicola GGPP synthase or
30   carotenoid biosynthesis enzyme structural gene, and is
free from a functional gene whose expression product
consumes the desired carotenoid, or otherwise inhibits
carotenoid production.

35

-24-

C. **G nes Encoding Enzymes for GGPP and Car tenoid**
   Biosynth sis

   1. **Isolation of the carotenoid g ne cluster**

   The plasmid pARC376 contains an approximately
5   13 kb chromosomal DNA fragment isolated by Perry et al.
   J. Bacteriol., 168:607 (1986) from the bacterium
   Erwinia herbicola EHO-10 (Escherichia vulneris; ATCC
   39368) that when transferred into the bacterium E. coli
   causes the E. coli cells to produce a yellow pigment.
10 Plasmid pARC376 was referred to by those authors as
   plasmid pPL376. A restriction map of the pARC376
   plasmid is shown in Figure 5.

   The structural genes in the plasmid
   responsible for pigment production are present on a DNA
15 fragment of about 7900 base pairs (bp) that is bounded
   by the restriction sites Pst I (at about position 4886)
   and Bgl II (at about position 12349) shown in Figure 5.
   There are a total of six relevant genes in this
   approximately 7900 bp region that cause the E. coli
20 cells to produce GGPP and the carotenoids phytoene
   through zeaxanthin diglucoside, which is the final
   product identified in the carotenoid pathway contained
   in plasmid pARC376.

   The biosynthetic pathway for carotenoid
25 production through zeaxanthin diglucoside is shown in
   Figure 1. E. coli cells, and all cells contemplated as
   hosts herein, naturally synthesize the isoprenoid
   intermediate farnesyl pyrophosphate (FPP). The genes
   for geranylgeranyl pyrophosphate (GGPP) synthase,
30 phytoene synthase, phytoene dehydrogenase-4H, lycopene
   cyclase, beta-carotene hydroxylase, and zeaxanthin
   glycosylase are located in the approximately 7900 bp
   DNA fragment in pARC376. E. coli cells that are
   transformed with the plasmid pARC376 are able to

convert some of the endogenous FPP into carotenoids by
utilizing the enzymes ncoded on the plasmid.

The following are descriptions of the
individual structural genes, including the genes of
5   this invention responsible for the synthesis of GGPP
and carotenoids, and the recombinant DNA manipulations
that have been performed to influence carotenoid
biosynthesis in bacteria such as E. coli, yeast such as
S. cerevisiae and higher plants.

10

2.   GGPP Synthase Gene and Plasmid Constructs
a.   DNA segments

An isolated, purified DNA segment comprising a
nucleotide sequence of at least 850 base pairs that
15   define a structural gene for the Erwinia herbicola
enzyme GGPP synthase and its DNA variants is one
structural gene contemplated by this invention. A
typical, useful DNA segment contains about 850 to about
1150 base pairs, whereas a more preferred DNA segment
20   contains about 850 to about 1000 base pairs. The
native sequence includes about 924 bp. Larger DNA
segments are also contemplated and are discussed
hereinafter.

An approximately 1153 bp fragment that extends
25   from the Bgl II (about 12349) site to the Eco RV (about
11196) site of plasmid pARC376 is shown in Figure 5. A
preferred structural gene for GGPP synthase is within
the about 1153 bp Bgl II to Eco RV restriction fragment
shown in Figure 5 and contains the previously mentioned
30   native structural gene of about 924 bp. This
structural gene is within the approximately 1030 bp Nco
I-Eco RV restriction fragment of plasmid pARC417BH.

Surprisingly it has been found that a
recombinant structural gene that encodes an amino-
35   terminal truncated version of this enzyme in which the

amino-terminal thirteen residues of the native enzyme
were deleted and were replaced by four extraneous amino
acid residues from the pARC306A vector was more active
(about two times) than was a recombinantly produced

5   enzyme having the encoded, native thirteen amino-
terminal residues. This more active enzyme is encoded
by the structural GGPP synthase gene containing about
1000 bp shown in Figure 3, and is within the
approximately 1150 bp segment Nco I-Pvu II restriction

10   fragment of plasmid pARC489B.
        Still more surprisingly, it has also been
found that truncation of the carboxy-terminus of the
GGPP synthase molecule made the enzyme still more
active. Thus, use of a GGPP synthase structural gene

15   of Figure 3 from which the 3' Bal I-Eco RV fragment was
removed provided the most active GGPP synthase found.
This structural gene of about 850 bp is within the
approximately 1000 bp Nco I-Pvu II restriction fragment
of pARC489D. This GGPP synthase gene is most preferred

20   herein. Details of the above work are described
hereinafter.
        The DNA sequence 1 from E. uredovora in
EP 0 393 690 is said there to encode the gene for
converting prephytoene pyrophosphate to phytoene. The DNA

25   sequence of that European application has about 59 percent
identity with the GGPP synthase illustrated herein. That
E. uredovora DNA sequence 1 is an analog of the before-
discussed GGPP synthase gene, and can also be used herein
for preparing GGPP.

30

        b.   Recombinant DNA molecules
        Also useful in this invention, are recombinant
DNA molecules comprising a vector operatively linked to
an exogenous DNA segment defining a structural gene

35   capable of expressing the enzyme GGPP synthase, as

-27-

described above, and a promoter suitable for driving
the expression of the gene encoding the enzyme in a
compatible host organism. The vector and promoter are
as described elsewhere herein. Particularly preferred
5   plasmid vectors include pARC417BH, pARC489B, pARC489D
and pARC145G.

     3.   **Phytoene Synthase Gene and Plasmid Construct**
          a.   DNA segments
10   An isolated, purified DNA segment comprising a
nucleotide sequence of at least about 927 base pairs
that define a structural gene for the Erwinia herbicola
enzyme phytoene synthase and its DNA variants is also
contemplated in this invention by producing phytoene
15   from GGPP. This structural gene typically contains
about 1000 to about 1250 bp including the 927 bp of the
native sequence, but can also contain a greater number
as discussed hereinafter. The structural gene for
phytoene synthase lies between positions 6383 and 5457
20   of plasmid pARC376 (Figure 5).
          A phytoene synthase gene useful herein at
least includes a sequence shown in Figure 4. In
preferred practice, the structural gene also includes
an upstream sequence shown in Figure 4 from about
25   position 8 (Bgl II site) to about position 15 (Nco I
site).
          A preferred phytoene synthase gene is within
the about 1112 bp Nco I-Eco RI fragment of plasmid
pARC285. Also included within that about 1112 bp
30   segment is the approximately 1040 bp Nco I-Bam HI
fragment that also encodes the desired structural gene.
          The most preferred structural gene includes a
nucleotide base sequence in Figure 4 from about base 8
to about base 15 as well as from about base 841 to
35   about base 1040, and contains about 1090 bp. This most

preferred gene is contained in the approximately 1176
base pair sequence of the Hpa I to Bam HI restriction
sites and approximately 1238 bp Pvu II-Eco RI fragments
present in the plasmid pARC140N, as well as in the
5   approximately 1088 bp sequence of the Bgl II-Eco RI
fragment of plasmid pARC140R.

b. Recombinant DNA molecules

A recombinant DNA molecule, comprising a
10   vector operatively linked to an exogenous DNA segment
defining a structural gene capable of expressing the
enzyme phytoene synthase, as discussed above, and a
promoter suitable for driving the expression of the
gene in a compatible host organism, is also useful in
15   this invention. The vector and promoter of this
recombinant molecule are also as are discussed herein.
Particularly preferred plasmid vectors include pARC285,
pARC140N, and pARC145G.

20   **4.   Phytoene Dehydrogenase-4H Gene and Plasmid
Construct**

a. DNA Segment

Another DNA segment of this invention is an
isolated DNA segment comprising a nucleotide sequence
25   that contains at least about 1470 bp that includes a
sequence defining a structural gene capable of
expressing the <u>Erwinia herbicola</u> enzyme phytoene
dehydrogenase-4H and its DNA variants. Phytoene
dehydrogenase-4H converts phytoene to lycopene. This
30   phytoene dehydrogenase-4H enzyme has a molecular mass
of about 51,000 daltons. The native phytoene
dehydrogenase-4H structural gene contains about 1470 bp
and is located between positions 7849 and 6380 of
plasmid pARC376.

A typical, useful DNA segment contains about 1500 base pairs and lies within the approximately 1891 bp Ava I (8231) to Nco I (6342) DNA fragment from pARC376 illustrated in Figure 5. Larger DNA segments are also contemplated, as discussed hereinafter.

A preferred DNA segment includes a nucleotide base sequence shown in Figure 11 from about base 15 to about base 1470. Particularly preferred DNA segments include the bases between the engineered Nco I site at about position 5 of Figure 11-1 (the initial Met residue) and about position 1470 of Figure 11-4, and is present in the approximately 1505 bp Nco I-Nco I restriction fragment (Nco I fragment) of plasmid pARC496A, the approximately 1508 bp Sal I-Sal I restriction fragment (Sal I fragment) of plasmid pARC146D, and the approximately 1506 bp Sph I-Nco I fragment present in plasmid pATC228. The sequence of the about 1508 bp Sal I fragment is illustrated in Figure 15.

A still further particularly preferred DNA segment is the approximately 2450 bp Xba I-Xba I fragment present in plasmid pATC1616. This fragment contains an approximately 1683 bp portion that encodes a chloroplast transit peptide of tobacco ribulose bis-phosphate carboxylase-oxygenase (hereinafter referred to as a chloroplast transit peptide) (about 177 bp) operatively linked in frame to the 5' end of the above Sph I-Nco I about 1506 bp phytoene dehydrogenase-4H gene. That approximately 1683 bp fragment is flanked at its 5' end by an about 450 bp CaMV 35S promoter sequence and at its 3' end by an about 300 bp NOS polyadenylation sequence.

This DNA segment can be used for expression of phytoene dehydrogenase-4H in higher plants and transport of the expressed phytoene dehydrogenase-4H

into chloroplasts such as those of tobacco. Infection
of a higher plant such as tobacco with <u>A. tumefaciens</u>
containing plasmid pATC1616 caused genomic
incorporation of DNA for the promoter, transit peptide-
phytoene dehydrogenase-4H and NOS sequence, and makes
the resultant plants resistant to the herbicide
norflurazon.

It is noted that restriction fragments having
the same restriction enzyme cleavage sequence at both
the 5' and 3' ends are sometimes referred to herein by
reference to a single restriction enzyme. Thus, the
approximately 1505 bp Nco I-Nco I restriction fragment
referred to above can also be referred to herein as an
approximately 1505 bp Nco I fragment. Similarly, the
approximately 1508 bp Sal I-Sal I fragment can be
referred to as the approximately 1508 bp Sal I
fragment, and the approximately 2450 bp Xba I-Xba I
fragment can be referred to as the approximately 2450
bp Xba I fragment.

b. Recombinant DNA Molecules

A recombinant DNA molecule comprising a vector
operatively linked to an exogenous DNA segment defining
a structural gene capable of expressing the enzyme
phytoene dehydrogenase-4H and a promoter suitable for
driving the expression of the enzyme in a compatible
host organism is also contemplated by this invention.
The structural gene has a nucleotide base sequence
described above. Particularly preferred plasmids
include pARC496A, pARC146D, pATC228 and pATC1616.

5. Lycopene Cyclase Gene and Plasmid Construct
a. DNA Segment

Also contemplated by this invention is an
isolated DNA segment comprising a nucleotide sequence

that contains at least about 1125 base pairs that
includes a sequence defining a structural gene capable
of expressing the Erwinia herbicola enzyme lycopene
cyclase and its DNA variants. This lycopene cyclase
enzyme has a molecular mass of about 39,000 daltons and
converts lycopene to β-carotene. A typical, useful DNA
segment contains at least about 1125 base pairs and
preferably at least about 1150 base pairs and lies
within the approximately 1548 bp Sal I (9340) to Pst I
(7792) DNA fragment from plasmid pARC376 illustrated in
Figure 5. The native Erwinia herbicola structural gene
for lycopene cyclase contains about 1125 base paris and
is located between positions 9002 and 7878 of plasmid
pARC376. Larger DNA segments are also contemplated, as
discussed hereinafter.

        A preferred DNA segment includes a nucleotide
base sequence shown in Figure 19, panels 1 and 2, from
about base 1 to about base 1222. A more preferred
sequence of about 1140 bp is present in the
approximately 1142 bp Sph I-Bam HI restriction fragment
of the plasmid pARC1509, and shown in Figure 19.

        A still further particularly preferred DNA
segment is an approximately 1319 bp Nco I-Bam HI
fragment. This fragment contains an approximately 177
bp portion that encodes a chloroplast transit peptide
operatively linked in frame to the 5' end of the above
Sph I-Bam HI 1142 bp lycopene cyclase gene. This DNA
segment can be used for expression of lycopene cyclase
in higher plants and transport of the expressed
lycopene cyclase into chloroplasts such as those of
tobacco.

        b. Recombinant DNA Molecules
        A recombinant DNA molecule comprising a vector
operatively linked to an exogenous DNA segment defining

-32-

a structural gene capable of expressing the enzyme
lycopene cyclase and a promoter suitable for driving
the expression of that enzyme in a compatible host
organism, is also contemplated by this invention. The
structural gene has a nucleotide base sequence
described above. Particularly preferred plasmid
vectors include pARC1510, pARC1520, pARC1509.

6.   Beta-Carotene Hydroxylase Gene and Plasmid Construct

a. DNA Segment

Further contemplated by this invention is an
isolated DNA segment comprising a nucleotide sequence
that contains at least about 531 base pairs and more
preferably about 878 base pairs, including a sequence
defining a structural gene capable of expressing the
Erwinia herbicola enzyme beta-carotene hydroxylase, an
enzyme that synthesizes zeaxanthin, and DNA variants of
that gene. The native enzyme is encoded between
positions 4991 and 5521 of plasmid pARC376.

One nucleotide base sequence corresponds to
the sequence in Figure 20 from about base 1 to about
base 894 at the Sma I site, and preferably from about
base 1 to about base 752. More preferably, the DNA
segment utilized is that shown in Figure 21 from about
base 25 to about base 897. The latter sequence
constitutes the about 870 bp Nco I to Sma I DNA
fragment contained in plasmid pARC406BH. A
contemplated DNA segment lies within the (4991) to
(5861) DNA segment of about 870 base pairs of plasmid
pARC376 illustrated in Figure 5.

A still further particularly preferred DNA
segment is an Xba I-Xba I fragment including about 1797
bp constituted by the following sequence of genes: (a)
the about 450 bp CaMV 35S promoter, (b) the 177 bp

-33-

sequence that encodes a chlorplast transit peptide
operatively linked in fram to the 5' end of (c) the
about 870 bp beta-carotene hydroxylase gene, and (d)
the about 300 bp NOS polyadenylation sequence. This

5   Xba I fragment can be cloned into the Xba I site of
plasmid pGA482, with the resulting plasmid being used
to transform A. tumefaciens. The resulting,
transformed A. tumefaciens can then be used to
transform higher plants such as tobacco wherein the

10   transit peptide-linked enzyme is expressed and
transporated to chlorplasts for the production of
zeaxanthin.

                 b.   **Recombinant DNA Molecules**

15   A recombinant DNA molecule comprising a vector
operatively linked to an exogenous DNA segment defining
a structural gene capable of expressing the Erwinia
herbicola enzyme beta-carotene hydroxylase and a
promoter suitable for driving the expression of that

20   enzyme in a compatible host organism, is also
contemplated by this invention. The structural gene
has a nucleotide base sequence described above.
Particularly preferred recombinant DNA molecules
include E. coli plasmid pARC404BH that contains the

25   about 874 bp Nco I-Sma I fragment, E. coli plasmid
pARC406BH that includes that same restriction fragment
driven by the Rec 7 promoter, and plasmid pARC145H
designed for S. cerevisiae expression of GAL 1- or GAL
10-driven or PGK-driven and URA 3-terminated beta-

30   carotene hydroxylase, as well as expression of GGPP
synthase and phytoene synthase driven by the GAL 1 and
GAL 10 promoters.

-34-

7.   Z axanthin Glycosylase Gene and Plasmid
     Construct
a.   DNA S gment
        Also contemplated by this invention is an

5   isolated DNA segment comprising a nucleotide sequence
    that contains at least about 1200 base pairs, including
    a sequence defining a structural gene capable of
    expressing the Erwinia herbicola enzyme zeaxanthin
    glycosylase, an enzyme that synthesizes zeaxanthin

10 diglucoside, and variant DNAs. The native DNA sequence
    for Erwinia herbicola lies between positions 10232 and
    9033 of plasmid pARC376. A preferred DNA encoding
    zeaxanthin glycosylase is in the about 1390 bp Nde I-
    Ava I fragment of plasmid pARC2019. The DNA sequence

15 of the E. herbicola structural gene is shown in Figure
    25 (SEQ ID NO:97).

        A further particularly preferred DNA segment
    is in an Xba I-Xba I fragment including about 2127 bp
    constituted by the following sequence of genes: (a) the

20 abouat 450 bp CaMV 35S promoter, (b) the 177 bp
    sequence that encodes a chloroplast transit peptide
    operatively linked in frame to the 5' end of (c) the
    about 1200 bp zeaxanthin glycosylase gene, and (d) the
    about 300 bp NOS polyadenylation sequence. This Xba I

25 fragment can be cloned into the Xba I site of plasmid
    pGA482, with the resulting plasmid being used to
    transform A. tumefaciens. The resulting, transformed
    A. tumefaciens can then be used to transform higher
    plants such as tobacco wherein the transit peptide-

30 linked enzyme is expressed and transporated to
    chlorplasts for the production of zeaxanthin
    diglucoside.

b. R combinant DNA Molecul s

A recombinant DNA molecule comprising a vector
operatively linked to an exogenous DNA segment defining
a structural gene capable of expressing the Erwinia
herbicola enzyme zeaxanthin glycosylase and a promoter
suitable for driving the expression of that enzyme in a
compatible host organism, is also contemplated by this
invention. The structural gene has a nucleotide base
sequence described above. Particularly preferred
recombinant DNA molecules include the E. coli plasmid
pARC2019 that contains the about 1390 bp Nde I-Ava I
fragment, an E. coli plasmid driven by the Rec 7
promoter, and a plasmid designed for S. cerevisiae
expression of a GAL-1-, GAL-10- or PGK-driven
zeaxanthin glycosylase.

8.   DNA Size

The previously described DNA segments are
noted as having a minimal length, as well as total
overall lengths. That minimal length defines the
length of a DNA segment having a sequence that encodes
a particular protein enzyme. Inasmuch as the coding
sequences for each of the six genes disclosed herein
are illustrated in the accompanying figures, isolated
DNA segments and variants thereof can be prepared by in
vitro mutagenesis, as described in the examples, that
begin at the initial ATG codon for a gene and end at or
just downstream of the stop codon for each gene. Thus,
a desired restriction site can be engineered at or
upstream of the initiation codon, and at or downstream
of the stop codon so that shorter structural genes than
most of those discussed above can be prepared, excised
and isolated.

As is well known in the art, so long as the
required DNA sequence is present, (including start and

stop signals), additional base pairs can be present at
either end of the segment and that segment can still be
utilized to express the protein. This, of course,
presumes the absence in the segment of an operatively

5   linked DNA sequence that represses expression,
expresses a further product that consumes an enzyme
desired to be expressed, expresses a product that
consumes a wanted reaction product produced by that
desired enzyme, or otherwise interferes with the

10   structural gene of the DNA segment.

Thus, so long as the DNA segment is free of
such interfering DNA sequences, a DNA segment of the
invention can be 2,000-15,000 base pairs in length.
The maximum size of a recombinant DNA molecule,

15   particularly an expression vector, is governed mostly
by convenience and the vector size that can be
accommodated by a host cell, once all of the minimal
DNA sequences required for replication and expression,
when desired, are present. Minimal vector sizes are

20   well known. Such long DNA segments are not preferred,
but can be used.

Example 4b illustrates that a DNA segment of
several thousand base pairs that contains the
structural genes for GGPP synthase and phytoene

25   synthase can be used to produce phytoene. The same
situation is true for phytoene dehydrogenase-4H
production as is seen in Example 9b. The DNA segment
used in Example 9b contains structural genes for GGPP
synthase, phytoene synthase and phytoene dehydrogenase-

30   4H, lycopene cyclase and the other structural genes for
zeaxanthin preparation. However, the gene for lycopene
cyclase, which utilizes lycopene, was impaired so that
no functional lycopene cyclase was produced and
lycopene accumulated. A similar situation is

35   illustrated in Example 16b wherein the gene for

β-carotene hydroxylase originally present in plasmid
pARC376 was made inoperative and β-carotene was found
to accumulate.

5   **9.   Construction of Plasmids**

        b.   DNA segments

            DNA segments that encode the before-described
enzyme proteins can be synthesized by chemical
techniques, for example, the phosphotriester method of
10   Matteucci et al., J. Am. Chem. Soc., 103:3185 (1981).
(The disclosures of the art cited herein are
incorporated herein by reference.) Of course, by
chemically synthesizing the coding sequence, any
desired modifications can be made simply by
15   substituting the appropriate bases for those encoding
the native amino acid residue sequence. However, DNA
segments including sequences discussed previously are
preferred.

            Furthermore, DNA segments containing
20   structural genes encoding the enzyme proteins can be
obtained from recombinant DNA molecules (plasmid
vectors) containing those genes. For instance, the
plasmid type recombinant DNA molecules pARC417BH,
pARC489B, pARC489D, pARC285, and pARC140N each contain
25   DNA sequences encoding different portions of the GGPP
synthase and phytoene synthase proteins and together
possess the entire sequence of DNA necessary for
expression of either protein in biologically active
form. Plasmid pARC145G contains DNA segments encoding
30   both enzymes. In addition, the plasmid type
recombinant DNA molecules pARC496A, pARC146D, pATC228
and pATC1616 each contain a DNA sequence encoding
biologically active phytoene dehydrogenase-4H proteins.
Similarly, the plasmid type recombinant DNA molecules
35   pARC1509, pARC1510, and pARC1520 each contain a DNA

sequence encoding biologically active lycopene cyclase
proteins. Also, the plasmid type recombinant DNA
molecules pARC404BH, pARC406BH and pARC145H each
contain a DNA sequence encoding biologically active

5   beta-carotene hydroxylase proteins, whereas plasmid
pARC2019 contains a DNA sequence that encodes
biologically active zeaxanthin glycosylase.
                    Plasmids pARC417BH, pARC489B, pARC489D,
pARC285, pARC140N and pARC145G have been deposited

10   pursuant to Budapest Treaty requirements with the
American Type Culture Collection (ATCC) 12301 Parklawn
Drive, Rockville, MD 20852 on February 26, 1990 and
were assigned the following respective accession
numbers 40755, 40758, 40757, 40756, 40759, and 40753.

15   Plasmids pARC496A, pARC146D and pATC228 were deposited
pursuant to Budapest Treaty requirements with the
American Type Culture Collection, (ATCC) 12301 Parklawn
Drive, Rockville, MD 20852 on May 11, 1990 and were
assigned the following respective accession numbers

20   40803, 40801 and 40802. Plasmid pATC1616 was similarly
deposited on May 15, 1990 and was assigned accession
No. 40806. Also, plasmids pARC1509, pARC1510, and
pARC1520 were deposited pursuant to Budapest Treaty
requirements with the American Type Culture Collection,

25   (AT) 12301 Parklawn Drive, Rockville, MD 20852 on July
27, 1990 and were assigned the following respective
accession numbers 40850, 40851 and 40852. Plasmids
pARC404BH, pARC406BH and pARC145H were similarly
deposited on January 16, 1991, and were assigned ATCC

30   accession numbers 40943, 40945, and 40944,
respectively. Plasmid pARC2019 was also deposited in
accordance with the Budapest Treaty on February 13,
1991 and received ATCC accession number 40974.
                    A DNA segment that includes a DNA sequence

35   encoding zeaxanthin glycosylase, beta-carotene

hydroxylase, lycopene cyclase, phytoene dehydrogenase-
4H, GGPP synthase, and phytoene synthase can be
prepared by excising and operatively linking
appropriate restriction fragments from each of the
5   above deposited plasmids using well known methods. The
DNA molecules of the present invention produced in this
manner typically have cohesive termini, i.e.,
"overhanging" single-stranded portions that extend
beyond the double-stranded portion of the molecule.
10   The presence of cohesive termini on the DNA molecules
of the present invention is preferred, although
molecules having blunt termini are also contemplated.
        Ribonucleic acid (RNA) equivalents of the
above described DNA segments are also contemplated.
15

        c. Recombinant DNA Molecules
        A recombinant DNA molecule of the present
invention can be produced by operatively linking a
vector to a DNA segment of the present invention to
20   form a plasmid such as those discussed and deposited
herein. Particularly preferred recombinant DNA
molecules are discussed in detail in the examples,
hereafter. Vectors capable of directing the expression
of GGPP synthase, phytoene synthase, phytoene
25   dehydrogenase-4H, lycopene cyclase, beta-carotene
hydroxylase, and/or zeaxanthin glycosylase genes are
referred to herein as "expression vectors".
        The expression vectors described above contain
expression control elements including the promoter.
30   The polypeptide coding genes are operatively linked to
the expression vector to allow the promoter sequence to
direct RNA polymerase binding and expression of the
desired polypeptide coding gene. Useful in expressing
the polypeptide coding gene are promoters which are
35   inducible, viral, synthetic, constitutive as described

by Poszkowski et al., EMBO J., 3:2719 (1989) and Odell
et al., Nature, 313:810 (1985), and temporally
regulated, spatially regulated, and spatiotemporally
regulated as given in Chua et al., Science, 244:174-181

5   (1989).

The choice of which expression vector and
ultimately to which promoter a polypeptide coding gene
is operatively linked depends directly on the
functional properties desired, e.g. the location and

10 timing of protein expression, and the host cell to be
transformed. These are well known limitations inherent
in the art of constructing recombinant DNA molecules.
However, a vector useful in practicing the present
invention is capable of directing the replication, and

15 preferably also the expression (for an expression
vector) of the polypeptide coding gene included in the
DNA segment to which it is operatively linked.

In one preferred embodiment, a vector includes
a prokaryotic replicon; i.e., a DNA sequence having the

20 ability to direct autonomous replication and
maintenance of the recombinant DNA molecule
extrachromosomally in a prokaryotic host cell
transformed therewith. Such replicons are well known
in the art.

25 Those vectors that include a prokaryotic
replicon can also include a prokaryotic promoter region
capable of directing the expression of the GGPP
synthase, phytoene synthase, phytoene dehydrogenase-4H,
lycopene cyclase, beta-carotene hydroxylase or

30 zeaxanthin glycosylase genes in a host cell, such as
E. coli, transformed therewith. Promoter sequences
compatible with bacterial hosts are typically provided
in plasmid vectors containing one or more convenient
restriction sites for insertion of a DNA segment of the

35 present invention. Typical of such vector plasmids are

-41-

pUC8, pUC9, and pBR329 available from Biorad
Laboratories, (Richmond, CA) and pPL and pKK223-3
available from Pharmacia, Piscataway, N.J.   A
particularly preferred promoter for use in prokaryotic
5   cells such as E. coli is the Rec 7 promoter present in
plasmid vectors pARC306A, pARC496A and pARC136, and
inducible by exogenously supplied nalidixic acid.
        Expression vectors compatible with eukaryotic
cells, preferably those compatible with yeast cells or
10   more preferably those compatible with cells of higher
plants, are also contemplated herein. Such expression
vectors can also be used to form the recombinant DNA
molecules of the present invention. Vectors for use in
yeasts such as S. cerevisiae can be episomal or
15   integrating, as is well known. Eukaryotic cell
expression vectors are well known in the art and are
available from several commercial sources.
        Normally, such vectors contain one or more
convenient restriction sites for insertion of the
20   desired DNA segment and promoter sequences. Exemplary
promoters for use in S. cerevisiae include the
S. cerevisiae phosphoglyceric acid kinase (PGK)
promoter and the divergent promoters GAL 10 and GAL 1.
        Typical vectors useful for expression of genes
25   in higher plants are well known in the art and include
vectors derived from the tumor-inducing (Ti) plasmid of
Agrobacterium tumefaciens described by Rogers et al.,
Meth. in Enzymol., 153:253-277 (1987). However,
several other expression vector systems are known to
30   function in plants including pCaMVCN transfer control
vector described by Fromm et al., Proc. Natl. Acad.
Sci. USA, 82:5824 (1985). Plasmid pCaMVCN (available
from Pharmacia, Piscataway, NJ) includes the
cauliflower mosaic virus CaMV 35S promoter. The

introduction of genes into higher plants is discussed
in greater detail hereinafter.

The use of retroviral expression vectors to
form the recombinant DNAs of the present invention is
5   also contemplated. As used herein, the term
"retroviral expression vector" refers to a DNA molecule
that includes a promoter sequence derived from the long
terminal repeat (LTR) region of a retrovirus genome.

Since some of these carotenoid products are to
10   be associated with food production and coloration, the
retroviral expression vector is preferably replication-
incompetent in eukaryotic cells. The construction and
use of retroviral vectors has been described by Verma,
PCT Publication No. WO87/00551, and Cocking et al,
15   Science, 236:1259-62 (1987).

In preferred embodiments, the vector used to
express the polypeptide coding gene includes a
selection marker that is effective in a plant cell,
preferably a drug resistance selection marker. One
20   preferred drug resistance marker is the gene whose
expression results in kanamycin resistance, i.e., the
chimeric gene containing the nopaline synthase
promoter, Tn5 neomycin phosphotransferase II and
nopaline synthase 3' nontranslated region described by
25   Rogers et al., in Methods For Plant Molecular Biology,
A. Weissbach and H. Weissbach, eds., Academic Press
Inc., San Diego, CA (1988). Another preferred marker
is the assayable chloramphenicol acetyltransferase
(cat) gene from the transposon Tn9.

30   A variety of methods has been developed to
operatively link DNA to vectors via complementary
cohesive termini or blunt ends. For instance,
complementary homopolymer tracts can be added to the
DNA segment to be inserted and to the vector DNA. The
35   vector and DNA segment are then joined by hydrogen

bonding between the complementary homopolymeric tails
to form recombinant DNA molecules.

Alternatively, synthetic linkers containing
one or more restriction endonuclease sites can be used
to join the DNA segment to the expression vector. The
synthetic linkers are attached to blunt-ended DNA
segments by incubating the blunt-ended DNA segments
with a large excess of synthetic linker molecules in
the presence of an enzyme that is able to catalyze the
ligation of blunt-ended DNA molecules, such as
bacteriophage T4 DNA ligase. Thus, the products of the
reaction are DNA segments carrying synthetic linker
sequences at their ends. These DNA segments are then
cleaved with the appropriate restriction endonuclease
and ligated into an expression vector that has been
cleaved with an enzyme that produces termini compatible
with those of the synthetic linker. Synthetic linkers
containing a variety of restriction endonuclease sites
are commercially available from a number of sources
including New England BioLabs, Beverly, MA.

Also contemplated by the present invention are
RNA equivalents of the above described recombinant DNA
molecules.

d. Introducing genes into higher plants

Methods for introducing polypeptide coding
genes into higher, multicelled plants include
<u>Agrobacterium</u>-mediated plant transformation, protoplast
transformation, gene transfer into pollen, injection
into reproductive organs and injection into immature
embryos. Each of these methods has distinct advantages
and disadvantages. Thus, one particular method of
introducing genes into a particular plant species may
not necessarily be the most effective for another plant

-44-

species, but it is well known which methods are useful
for a particular plant species.

Agrobacterium-mediated transfer is a widely
applicable system for introducing genes into plant
5   cells because the DNA can be introduced into whole
plant tissues, thereby bypassing the need for
regeneration of an intact plant from a protoplast. The
use of Agrobacterium-mediated expression vectors to
introduce DNA into plant cells is well known in the
10   art. See, for example, the methods described by Fraley
et al., Biotechnology, 3:629 (1985) and Rogers et al.,
Methods in Enzymology, 153:253-277 (1987). Further,
the integration of the Ti-DNA is a relatively precise
process resulting in few rearrangements. The region of
15   DNA to be transferred is defined by the border
sequences, and intervening DNA is usually inserted into
the plant genome as described by Spielmann et al., Mol.
Gen. Genet., 205:34 (1986) and Jorgensen et al., Mol.
Gen. Genet., 207:471 (1987).

20   Modern Agrobacterium transformation vectors
are capable of replication in E. coli as well as
Agrobacterium, allowing for convenient manipulations as
described by Klee et al., in Plant DNA Infectious
Agents, T. Hohn and J. Schell, eds., Springer-Verlag,
25   New York (1985) pp. 179-203.

Moreover, recent technological advances in
vectors for Agrobacterium-mediated gene transfer have
improved the arrangement of genes and restriction sites
in the vectors to facilitate construction of vectors
30   capable of expressing various polypeptide coding genes.
The vectors described by Rogers et al., Methods in
Enzymology, 153:253 (1987), have convenient multi-
linker regions flanked by a promoter and a
polyadenylation site for direct expression of inserted

polypeptide coding genes and are suitable for present
purposes.

In those plant species where _Agrobacterium_-
mediated transformation is efficient, it is the method
of choice because of the facile and defined nature of
the gene transfer. However, few monocots appear to be
natural hosts for _Agrobacterium_, although transgenic
plants have been produced in asparagus using
Agrobacterium vectors as described by Bytebier et al.,
Proc. Natl. Acad. Sci. U.S.A., 84:5345 (1987).
Therefore, commercially important cereal grains such as
rice, corn, and wheat must be transformed using
alternative methods.

Higher plants have the ability to produce
carotenoids. The site of synthesis for all plant
carotenoids is in the chloroplast. Carotenoid
biosynthesis is highly regulated in plants. Masoner
et al., _Planta_ 105:267 (1972); Frosch et al., _Planta_
148:279 (1980); Mohr, _Photosynthesis V. Chloroplast_
Development, pp. 869-883 (1981); Oelmueller et al.,
Planta 164:390 (1985); Harpster et al., _Physiol. Plant._
64:147 (1985); Steinmueller et al., _Molecular Form and_
Function of the Plant Genome, pp. 277-290 (1986).
Therefore, the ability to use recombinant DNA
technology to increase endogenous carotenoid
biosynthesis is questionable unless a novel approach is
used. However, using the genes for GGPP synthase,
phytoene synthase, phytoene dehydrogenase-4H, lycopene
cyclase, beta-carotene hydroxylase and zeaxanthin
glycosylase to induce zeaxanthin and zeaxanthin
diglucoside synthesis in the cytoplasm is a viable
approach, even though carotenoids are not naturally
produced in the cytoplasm.

Agrobacterium-mediated transformation of leaf
disks and other tissues appears to be limited to plant

species that _Agrobacterium_ naturally infects. Thus,
Agrobacterium-mediated transformation is most efficient
in dicotyledonous plants. However, as mentioned above,
the transformation of asparagus using _Agrobacterium_ can

5   also be achieved. See, for example, Bytebier, et al.,
Proc. Natl. Acad. Sci., 84:5345 (1987).
         Transformation of plant protoplasts can be
achieved using methods based on calcium phosphate
precipitation, polyethylene glycol treatment,

10   electroporation, and combinations of these treatments.
See, for example, Potrykus et al., _Mol. Gen. Genet._,
199:183 (1985); Lorz et al., _Mol. Gen. Genet._, 199:178
(1985); Fromm et al., _Nature_, 319:791 (1986); Uchimiya
et al., _Mol. Gen. Genet._, 204:204 (1986); Callis et

15   al., _Genes and Development_, 1:1183 (1987); and Marcotte
et al., _Nature_, 335:454 (1988).
         Application of these systems to different
plant species depends upon the ability to regenerate
that particular plant species from protoplasts.

20   Illustrative methods for the regeneration of cereals
from protoplasts are described in Fujimura et al.,
Plant Tissue Culture Letters, 2:74 (1985); Toriyama et
al., _Theor. Appl. Genet._, 73:16 (1986); Yamada et al.,
Plant Cell Rep., 4:85 (1986); Abdullah et al.,

25   _Biotechnology_, 4:1087 (1986).
         To transform plant species that cannot be
successfully regenerated from protoplasts, other ways
to introduce DNA into intact cells or tissues can be
utilized. For example, regeneration of cereals from

30   immature embryos or explants can be effected as
described by Vasil, _Biotechnology_, 6:397 (1988). In
addition, "particle gun" or high-velocity
microprojectile technology can be utilized. Using such
technology, DNA is carried through the cell wall and

35   into the cytoplasm on the surface of small metal

-47-

particles as described in Klein et al., Nature, 327:70
(1987); Klein et al., Proc. Natl. Acad. Sci. U.S.A.,
85:8502 (1988); and McCabe et al., Biotechnology, 6:923
(1988). The metal particles penetrate through several
5   layers of cells and thus allow the transformation of
cells within tissue explants.

Metal particles have been used to successfully
transform corn cells and to produce fertile, stably
transformed tobacco and soybean plants. Transformation
10   of tissue explants eliminates the need for passage
through a protoplast stage and thus speeds the
production of transgenic plants.

DNA can also be introduced into plants by
direct DNA transfer into pollen as described by Zhou et
15   al., Methods in Enzymology, 101:433 (1983); D. Hess,
Intern Rev. Cytol., 107:367 (1987); Luo et al., Plant
Mol. Biol. Reporter, 6:165 (1988). Expression of
polypeptide coding genes can be obtained by injection
of the DNA into reproductive organs of a plant as
20   described by Pena et al., Nature, 325:274 (1987). DNA
can also be injected directly into the cells of
immature embryos and the rehydration of desiccated
embryos as described by Neuhaus et al., Theor. Appl.
Genet., 75:30 (1987); and Benbrook et al., in
25   Proceedings Bio Expo 1986, Butterworth, Stoneham, MA,
pp. 27-54 (1986).

The regeneration of plants from either single
plant protoplasts or various explants is well known in
the art. See, for example, Methods for Plant Molecular
30   Biology, A. Weissbach and H. Weissbach, eds., Academic
Press, Inc., San Diego, CA (1988). This regeneration
and growth process includes the steps of selection of
transformant cells and shoots, rooting the transformant
shoots and growth of the plantlets in soil.

-48-

        The regeneration of plants containing the
foreign gene introduced by Agrobacterium from leaf
explants can be achieved as described by Horsch et al.,
Science, 227:1229-1231 (1985). In this procedure,
5   transformants are grown in the presence of a selection
agent and in a medium that induces the regeneration of
shoots in the plant species being transformed as
described by Fraley et al.,Proc. Natl. Acad. Sci.
U.S.A., 80:4803 (1983).
10   This procedure typically produces shoots
within two to four weeks and these transformant shoots
are then transferred to an appropriate root-inducing
medium containing the selective agent and an antibiotic
to prevent bacterial growth. Transformant shoots that
15   rooted in the presence of the selective agent to form
plantlets are then transplanted to soil or other media
to allow the production of roots. These procedures
vary depending upon the particular plant species
employed, such variations being well known in the art.
20   In higher plants, the transformed elements are
so manipulated as to permit them to mature into soil-
or otherwise-cultivated plants, such as plants that are
cultivated hydroponically or in other soil-free media
such as lava rock, crushed coral, sphagnum moss and the
25   like.
        Methods not utilizing tissue culture
procedures are also contemplated, for example, using
Agrobacterium-mediated vectors to produce transgenic
plants from seeds.
30   A plant of the present invention containing
one or more of the desired six enzyme proteins; i.e.,
GGPP synthase, phytoene synthase, phytoene
dehydrogenase-4H, lycopene cyclase, β-carotene
hydroxylase and zeaxanthin glycosylase, is cultivated
35   using methods well known to one skilled in the art.

Any of the transgenic plants of the pres nt invention
can be cultivated to isolate the desired carotenoid
products they contain.

After cultivation, the transgenic plant is
harvested to recover the carotenoid product. This
harvesting step can consist of harvesting the entire
plant, or only the leaves, or roots of the plant. This
step can either kill the plant or if only a non-
essential portion of the transgenic plant is harvested
can permit the remainder of the plant to continue to
grow.

In preferred embodiments this harvesting step
further comprises the steps of:

(i) homogenizing at least a carotenoid-
containing portion of the transgenic plant to produce a
plant pulp and using the carotenoid-containing pulp
directly, as in dried pellets or tablets as where an
animal food is contemplated; or

(ii) extracting the carotenoid(s) from the
plant pulp with an appropriate solvent such as an
organic solvent or by supercritical extraction [Favati
et al., J. Food Sci., 53:1532 (1988) and the citations
therein] to produce a carotenoid-containing liquid
solution or suspension; and

(iii) isolating the carotenoid(s) from the
solution or suspension.

At least a portion of the transgenic plant is
homogenized to produce a plant pulp using methods well
known to one skilled in the art. This homogenization
can be done manually, by a machine, or by a chemical
means as long as the transgenic plant portions are
broken up into small pieces to produce a plant pulp.
This plant pulp consists of a mixture of the carotenoid
of interest residual amounts of precursors, cellular
particles and cytosol contents. This pulp can be dried

and compressed into pellets or tablets and eaten or
otherwise used to derive the benefits, or the pulp can
be subjected to extraction procedures.

5   The carotenoid can be extracted from the
plant pulp produced above to form a solution or
suspension. Such extraction processes are common and
well known to one skilled in this art. For example,
the extracting step can consist of soaking or immersing
the plant pulp in a suitable solvent. This suitable

10   solvent is capable of dissolving or suspending the
carotenoid present in the plant pulp to produce a
carotenoid-containing solution or suspension. Solvents
useful for such an extraction process are well known to
those skilled in the art and include water, several

15   organic solvents and combinations thereof such as
methanol, ethanol, isopropanol, acetone, acetonitrile,
tetrahydrofuran (THF), hexane, and chloroform. A
vegetable oil such as peanut, corn, soybean and similar
oils can also be used for this extraction.

20   Isolation (harvesting) of carotenoids from
bacteria, yeasts, fungi and other lower organisms is
illustrated hereinafter using A. tumefaciens and
E. coli. Broadly, cells transfected with structural
genes for GGPP synthase, phytoene synthase, phytoene

25   dehydrogenase-4H, lycopene cyclase, beta-carotene
hydroxylase and zeaxanthin glycosylase, as desired, are
grown under suitable conditions for a period of time
sufficient for a desired carotenoid to be synthesized.
The carotenoid-containing cells, preferably in dried

30   form, are then lysed chemically or mechanically, and
the carotenoid is extracted from the lysed cells using
a liquid organic solvent, as described before, to form
a carotenoid-containing liquid solution or suspension.
The carotenoid is thereafter isolated from the liquid

solution or suspension by usual means such as
chromatography.

5   The carotenoid is isolated from the solution
or suspension produced above using methods that are
well known to those skilled in the art of carotenoid
isolation. These methods include, but are not limited
to, purification procedures based on solubility in
various liquid media, chromatographic techniques such
as column chromatography and the like.

10

    D.   **Methods for Preparing Caroentoid Biosynthesis**
       Enzymes
      1.   **Introduction**
         a.   **Transformed Cells and Cultures**

15   The present invention also relates to host
cells transformed with recombinant DNA molecules of the
present invention, preferably recombinant DNA capable
of expressing GGPP synthase and membrane-bound (or
soluble) phytoene synthase, phytoene dehydrogenase-4H,
20 lycopene cyclase, beta-carotene hydroxylase and
zeaxanthin glycosylase enzymes. These six enzymes can
be referred to as carotenoid biosynthesis enzymes.

      The host cells can be either prokaryotic or
eukaryotic. Bacterial cells are preferred prokaryotic
25 host cells and typically are a strain of E. coli such
as, for example the E. coli strain HB101, available
from BRL Life Technologies, Inc., Gaithersburg, MD
(BRL). Preferred eukaryotic host cells include yeast
and plant cells or protoplasts, preferably cells from
30 higher plants. Preferred eukaryotic host cells include
S. cerevisiae cells such as YPH499 obtained from
Dr. Phillip Hieter, Johns Hopkins University,
Baltimore, MD, discussed in Example 5.

      Transformation of appropriate cell hosts with
35 a recombinant DNA molecule of the present invention is

-52-

accomplished by well known methods that typically
depend on the type of vector used. With regard to
transformation of prokaryotic host cells, see, for
example, Cohen et al., <u>Proc. Natl. Acad. Sci. USA</u>,
5   69:2110 (1972); and Maniatis et al., <u>Molecular Cloning,</u>
<u>A Laboratory Manual</u>, Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY (1982). With regard to
transformation of plant cells with retroviral vectors
containing recombinant DNAs, see, for example, Verma,
10   PCT Publication No. WO 87/00551, 1987, who isolated
protoplasts from plant tissue, and inserted the
retroviral genome in proviral (double stranded) form
into the genome of the protoplasts. The transformed
protoplasts were developed into callus tissue and then
15   regenerated into transgenic plants. Plants derived
from the protoplasts and their progeny carry the
genetic material of the recombinant retroviral vector
in their genomes and express the protein product.
        Successfully transformed cells; i.e., cells
20   that contain a recombinant DNA molecule of the present
invention, can be identified by well known techniques.
For example, cells resulting from the introduction of a
recombinant DNA of the present invention can be cloned
to produce monoclonal colonies. Cells from those
25   colonies can be harvested, lysed and their DNA content
examined for the presence of the recombinant DNA using
a method such as that described by Southern, <u>J. Mol.</u>
<u>Biol.</u>, 98:503 (1975) or Berent et al., <u>Biotech.</u>, 3:208
(1985).
30   In addition to directly assaying for the
presence of recombinant DNA, successful transformation
can be confirmed by well known immunological methods
when the recombinant DNA is capable of directing the
expression of specific protein antigens. For example,
35   cells successfully transformed with an expression

v ctor may produce proteins displaying GGPP synthase,
phytoene synthase, phytoene dehydrogenase-4H, lycopene
cyclase, beta-carotene hydroxylase or zeaxanthin
glycosylase antigenicity.

Identifying successful transformation of E.
coli in this invention is relatively easy for
carotenoids, except phytoene. Carotenoid-containing
colonies formed are usually characterized by colored
pigment formation. For example, beta-carotene,
zeaxanthin and zeaxanthin diglucoside are yellow and
lycopene is red.

b. Methods for Producing Enzymes

A method is contemplated by this invention for
preparing a carotenoid biosynthesis enzyme. This
method comprises initiating a culture, in a nutrient
medium, of transformed prokaryotic or eukaryotic host
cells. The host cells are transformed with a
recombinant DNA molecule containing a compatible
expression vector operatively linked to a before-
described exogenous DNA segment that defines the
structural gene for a carotenoid biosynthesis enzyme,
as desired.

This invention further comprises cultures
maintained for a time period sufficient for the host
cells to express the carotenoid biosynthesis enzyme
protein molecules, which proteins can be recovered in
purified form if desired. Nutrient media useful for
culturing transformed host cells are well known in the
art and can be obtained from several commercial
sources. A further discussion of useful host cells
and nutrient media are provided in the following
section.

A further aspect contemplated is a method for
preparing one carotenoid biosynthesis enzyme in the

-54-

presence of either or all of the other carotenoid
biosynthesis enzymes such as GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H, lycopene cyclase,
β-carotene hydroxylase and zeaxanthin glycosylase.

5   This method is substantially identical to the before-
described method except that the host cells are also
transformed with a compatible expression vector
operatively linked to a before-described exogenous DNA
segment that defines any or all of the structural genes

10   for GGPP synthase, phytoene synthase, phytoene
dehydrogenase-4H, lycopene cyclase, β-carotene
hydroxylase and zeaxanthin glycosylase.

The transformed host cell can contain a single
expression vector that contains one or more of the six

15   structural genes. The host can also be transformed
with two expression vectors each containing a
structural gene for one or more of the enzymes; e.g.,
one for at least beta-carotene hydroxylase or
zeaxanthin glycosylase and another that contains at

20   least one of the other four (or five) enzymes, or three
expression vectors; e.g., one for at least beta-
carotene hydroxylase or zeaxanthin glycosylase, and two
others that each contain at least one of the other four
(or five) enzymes. An exemplary four expression vector

25   transformation can also be used in which at least one
expression vector contains the structural gene that
encodes beta-carotene hydroxylase or zeaxanthin
glycosylase, with each of the other vectors containing
at least one structural gene for each of the other

30   structural genes. A host cell can also be transformed
with five vectors; i.e., one expression vector that
contains the gene encoding each one of the first four
enzymes and another vector containing the last two
genes. A six-vector system can also be utilized for

35   production of zeaxanthin glycosylase in which a host is

transformed with one expression vector for each of the
six named enzymes.

2.   Methods for Preparing Carotenoids

A carotenoid can be produced by a method that
includes initiating a culture, in a nutrient medium, of
prokaryotic or eukaryotic host cells that are
transformed with a recombinant DNA molecule containing
a compatible expression vector operatively linked to a
before-described exogenous DNA segment that defines the
structural gene for a carotenoid biosynthesis enzyme
that converts an immediate precursor substrate molecule
into the desired carotenoid, and which cells provide
the immediate precursor molecule that is the substrate
for the expressed carotenoid biosynthesis enzyme. The
cell culture is maintained for a time period sufficient
for the transformed cells to produce (express) the
desired carotenoid biosynthesis enzyme, and for that
expressed enzyme to convert the provided immediate
precursor substrate molecule into the desired
carotenoid. The produced carotenoid can thereafter be
recovered as discussed herein. In higher plants, the
nutrient medium (and in many cases the enzyme substrate
that is the immediate precursor molecule) is supplied
by the plant itself, and the initiated culture is the
germinated seed, protoplast or even a grafted explant
from a prior culture.

In one embodiment where the host cells do not
themselves produce carotenoid biosynthesis enzymes or
immediate precursor substrate molecules, the required
carotenoid biosynthesis enzymes are provided to the
cells by transformation of those cells with one or more
exogenous recombinant DNA molecules that encode and
express the appropriate genes so that appropriate
enzymes and precursor substrate molecules are provided

to the cells. For example, in E. coli, an exogenous
structural gene for GGPP synthase is required to
produce GGPP from the ubiquitous precursor FPP, and an
exogenous structural gene for phytoene synthase is

5   required to convert GGPP (the immediate precursor
substrate) into phytoene. Thus, at least two exogenous
recombinant DNA molecules are needed to produce a
carotenoid.

        If the next carotenoid shown in Figure 1,

10   lycopene, is desired, the above transformed cells must
also be further transformed with an exogenous
recombinant DNA molecule that codes for and expresses
phytoene dehydrogenase-4H. Further transformation of
such transformed cells is needed as each ensuing

15   carotenoid illustrated in Figure 1 that is desired to
be prepared. Thus, with transformation with the six
structural genes discussed herein is accomplished for
the production of zeaxanthin diglucoside. The
exogenous structural genes used for the transformation

20   can reside in a single recombinant DNA molecule, or in
a plurality of such recombinant molecules as is
exemplified below.

        In one aspect for producing phytoene, the
recombinant DNA molecule contains an expression system

25   that comprises one or more expression vectors
compatible with host cells operatively linked to an
exogenous DNA segment that comprises (i) a nucleotide
base sequence corresponding to a sequence defining a
structural gene for GGPP synthase as discussed before,

30   and (ii) a nucleotide base sequence corresponding to a
sequence defining a structural gene for phytoene
synthase as also discussed before. A particularly
preferred expression vector plasmid pARC145G contains
structural genes for both GGPP synthase and phytoene

35   synthase, and produces phytoene in S. cerevisiae.

In another aspect for producing lycopene, the
recombinant DNA molecule pref rably contains an
expression system that comprises one or more expression
vectors compatible with host cells, operatively linked

5   to an exogenous DNA segment that comprises (i) a
nucleotide base sequence corresponding to a sequence,
defining a structural gene for GGPP synthase, and (ii)
a nucleotide base sequence corresponding to a sequence
defining a structural gene for phytoene synthase, and

10   (iii) a nucleotide base sequence corresponding to the
sequence defining a structural gene for phytoene
dehydrogenase-4H. Thus, phytoene is provided to the
host cells by the enzymes expressed by the expression
system.

15   In one particularly preferred aspect, the
structural genes for GGPP synthase, phytoene synthase,
and phytoene dehydrogenase-4H are contained operatively
linked in a single expression vector, preferably under
the control of the same promoter. In another preferred

20   aspect, two expression vectors are used, with the
structural genes for GGPP synthase and phytoene
synthase on one vector and the structural gene for
phytoene dehydrogenase-4H on the other vector. In yet
another preferred embodiment, three expression vectors

25   are used. Yeast and plants require a separate promoter
for each gene, although the same promoter can be used
for each gene.
         Example 9b illustrates lycopene production in
E. coli host cells using a single expression vector

30   (pARC376-Ava 102) containing all three genes.
Similarly, the very active GGPP synthase gene contained
in plasmid pARC489D and phytoene synthase gene
contained in plasmid pARC140N can be transformed
separately or together with the phytoene dehydrogenase-

35   4H structural gene found in plasmid pARC496A to prepare

transformed host <u>E. coli</u> cells that contain all three
functional structural g nes. Here, expression of
plasmids pARC489D and pARC140N provides the enzymes
needed to convert ubiquitous cellular precursors into
5   the required phytoene that is converted into lycopene
by the action of the phytoene dehydrogenase-4H
expressed by plasmid pARC496A. Likewise, Example 10
illustrates lycopene production in <u>S. cerevisiae</u> host
cells transformed with both plasmid pARC145G, whose
10   expression products provides phytoene to the cells, and
plasmid pARC146D that expresses phytoene dehydrogenase-
4H that converts the provided phytoene into lycopene.

In yet another aspect for the production of
β-carotene, the recombinant DNA molecule preferably
15   contains an expression system that comprises one or
more expression vectors compatible with host cells,
operatively linked to an exogenous DNA segment that
comprises (i) a nucleotide base sequence corresponding
to a sequence, defining a structural gene for GGPP
20   synthase, and (ii) a nucleotide base sequence
corresponding to a sequence defining a structural gene
for phytoene synthase, (iii) a nucleotide base sequence
corresponding to the sequence defining a structural
gene for phytoene dehydrogenase-4H, and (iv) a
25   nucleotide base sequence corresponding to the sequence
defining a structural gene for lycopene cyclase. Thus,
lycopene is provided to the host cells by the enzymes
expressed by the expression system.

In one particularly preferred aspect, the
30   structural genes for GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H and lycopene cyclase are
contained operatively linked in a single expression
vector, preferably under the control of the same
promoter. In another preferred aspect, two expression
35   vectors are used, with the structural genes for GGPP

-59-

synthase, phytoene synthase and phytoene dehydrogenase-4H on one vector and the structural gene for lycopene cyclase on the other vector.   In yet another preferred aspect, three expression vectors are used.   Yeast and
5   plants require a separate promoter for each gene, although the same promoter can be used for each gene.

      Example 16 illustrates beta-carotene production in E. coli host cells using a single expression vector plasmid pARC376-Pst 102 containing
10   all four genes.   Similarly, the very active GGPP synthase gene contained in plasmid pARC489D, phytoene synthase gene contained in plasmid pARC140N and the phytoene dehydrogenase-4H structural gene found in plasmid pARC496A can be transformed separately or
15   together with the lycopene cyclase structural gene found in plasmid pARC1510 to prepare transformed host E. coli cells that contain all four functional structural genes.   Here, expression of plasmids pARC489D, pARC140N and pARC496A provides the enzymes
20   needed to convert ubiquitous cellular precursors into the required phytoene that is converted into lycopene that is subsequently converted into beta-carotene by the action of the lycopene cyclase expressed by plasmid pARC1510.   Likewise, Example 17 illustrates beta-
25   carotene production in plasmid pARC145G, whose expression products provides phytoene to the cells and plasmid pARC1520 that expresses both phytoene dehydrogenase-4H, which converts the provided phytoene into lycopene, and lycopene cyclase that converts
30   lycopene into beta-carotene.

      In a still further aspect, the recombinant DNA molecule preferably contains an expression system that comprises one or more expression vectors compatible with host cells, operatively linked to an exogenous DNA
35   segment that comprises (i) a nucleotide base sequence

corresponding to a sequence, defining a structural gene
for GGPP synthase, and (ii) a nucleotide bas sequence
corresponding to a sequence defining a structural gene
for phytoene synthase, (iii) a nucleotide base sequence

5   corresponding to the sequence defining a structural
gene for phytoene dehydrogenase-4H, (iv) a nucleotide
base sequence corresponding to the sequence defining a
structural gene for lycopene cyclase and (v) a
nucleotide base sequence corresponding to the sequence

10   defining a structural gene for beta-carotene
hydroxylase. Thus, beta-carotene is provided to the
host cells by the enzymes expressed by the expression
system.

        In one particularly preferred aspect, the
15   structural genes for GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H, lycopene cyclase and beta-
carotene hydroxylase are contained operatively linked
in a single expression vector, preferably under the
control of the same promoter. In another preferred

20   aspect, two expression vectors are used, with the
structural genes for GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H and lycopene cyclase on one
vector and the structural gene for beta-carotene
hydroxylase on the other vector. In yet another

25   preferred aspect, three expression vectors are used.
Yeast and plants require a separate promoter for each
gene, although the same promoter can be used for each
gene.

        Example 1 illustrates zeaxanthin production in
30   E. coli and A. tumefaciens host cells using a single
expression plasmid vector pARC288 containing all five
genes. Example 21 illustrates zeaxanthin production in
E. coli host cells using an expression plasmid vector
pARC279 containing all four genes required to produce

35   beta-carotene, but with the gene for beta-carotene

hydroxylase having been deleted. E. coli containing
plasmid pARC279 were further transformed with the
plasmid pARC406BH, and then the cells were grown in
appropriate selective medium. Zeaxanthin was produced.

5   Similarly, the very active GGPP synthase gene
contained in plasmid pARC489D, the phytoene synthase
gene contained in plasmid pARC140N, the phytoene
dehydrogenase-4H gene found in plasmid pARC496A, the
lycopene cyclase structural gene found in plasmid

10   pARC1510 can be transformed separately or together with
the beta-carotene hydroxylase structural gene found in
plasmid pARC406BH to prepare transformed host E. coli
cells that contain all five functional structural
genes. Here, expression of plasmids pARC489D,

15   pARC140N, pARC496A and pARC1510 provides the enzymes
needed to convert ubiquitous cellular precursors into
the required phytoene that is converted into lycopene
and then beta-carotene, which is subsequently converted
into zeaxanthin by the action of the beta-carotene

20   hydroxylase expressed by plasmid pARC406BH. Likewise,
Example 22 illustrates zeaxanthin production in
S. cerevisiae host cells transformed with plasmid
pARC145H, whose expression products provide GGPP
synthase, phytoene synthase and beta-carotene

25   hydroxylase to the cells, and plasmid pARC1520 that
expresses both phytoene dehydrogenase-4H and lycopene
cyclase. Thus all enzymes required for zeaxanthin
biosynthesis are found on these two plasmids.

          In yet another aspect for preparing zeaxanthin

30   diglucoside, the recombinant DNA molecule preferably
contains an expression system that comprises one or
more expression vectors compatible with host cells,
operatively linked to an exogenous DNA segment,
comprising (i) a nucleotide base sequence corresponding

35   to a sequence, defining a structural gene for GGPP

-62-

synthase, and (ii) a nucleotide base sequence
corresponding to a sequence defining a structural gene
for phytoene synthase, (iii) a nucleotide base sequence
corresponding to the sequence defining a structural
5   gene for phytoene dehydrogenase-4H, (iv) a nucleotide
base sequence corresponding to the sequence defining a
structural gene for lycopene cyclase, (v) a nucleotide
base sequence corresponding to a sequence defining a
structural gene for beta-carotene hydroxylase, and a
10   nucleotide base sequence corresponding to a sequence
defining a structural gene for zeaxanthin glycosylase.
Thus, zeaxanthin is provided to the host cells by the
enzymes expressed by the expression system.

In one particularly preferred aspect, the
15   structural genes for GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H, lycopene cyclase, beta-
carotene hydroxylase and zeaxanthin glycosylase are
contained operatively linked in a single expression
vector, preferably under the control of the same
20   promoter. In another preferred aspect, two expression
vectors are used, with the structural genes for GGPP
synthase, phytoene synthase, phytoene dehydrogenase-4H,
lycopene cyclase and beta-carotene hydroxylase on one
vector and the structural gene for zeaxanthin
25   glycosylase on the other vector. In yet other
preferred aspects, three, four, five or six expression
vectors are used. Yeast and plants require a separate
promoter for each gene, although the same promoter can
be used for each gene.

30   Example 1 illustrates zeaxanthin diglucoside
production in and recovery from E. coli using a single
expression vector pARC376 containing all six genes.
Example 26 illustrates zeaxanthin diglucoside
production in E. coli host cells using expression
35   vector pARC288 containing all five genes required to

produce zeaxanthin plus plasmid pARC2019 with the gene for zeaxanthin glycosylase. Thus, <u>E. coli</u> containing pARC288 were further transformed with the plasmid pARC2019, and then the cells were grown in appropriate selective medium. Zeaxanthin diglucoside was produced and recovered.

Similarly, the very active GGPP synthase gene contained in plasmid pARC489D, the phytoene synthase gene contained in plasmid pARC140N, the phytoene dehydrogenase-4H gene found in plasmid pARC496A, lycopene cyclase structural gene found in plasmid pARC1510 and the beta-carotene hydroxylase structural gene found in plasmid pARC406GH can be transformed separately or together with the zeaxanthin glycosylase gene found in plasmid pARC2019 to prepare transformed host <u>E. coli</u> cells that contain all six functional structural genes. Here, expression of plasmids pARC489D, pARC140N, pARC496A, pARC1510 and pARC406BH provides the enzymes needed to convert ubiquitous cellular precursors into the required phytoene that is converted into lycopene and then beta-carotene, that is converted into zeaxanthin and then into zeaxanthin diglucoside by the action of the zeaxanthin glycosylase expressed by plasmid pARC406BH. Likewise, Example 27 illustrates zeaxanthin diglucoside production in <u>S.</u> <u>cerevisiae</u> host cells multiply transformed with plasmid pARC145H, whose expression products provide GGPP synthase, phytoene synthase and beta-carotene hydroxylase to the cells, plasmid pARC1520 that expresses both phytoene dehydrogenase-4H and lycopene cyclase, and a plasmid that expresses zeaxanthin glycosylase. Thus all enzymes required for zeaxanthin diglucoside biosynthesis are found on these three plasmids.

-64-

The order of expression of the structural
genes is not important so, for example, the structural
gene for GGPP synthase can be located 5' (upstream)
from the structural gene for phytoene synthase, or vice
5   versa.

As another embodiment, this method also
contemplates carotenoid production by use of
transformed host cells containing only one carotenoid
synthesis gene-containing expression vector. Here, the
10   nutrient medium supplies the immediate precursor
substrate molecule to the host cells so that those host
cells can provide the precursor substrate for the
expressed enzyme. The nutrient medium can contain the
requisite amount of precursor in micelles or vesicles,
15   as are well known, which are taken up by the host
cells.

Another aspect of this embodiment contemplates
host cells transformed with one, two, three, four or
five expression vectors for the production of phytoene
20   synthase, phytoene dehydrogenase-4H, lycopene cyclase,
beta-carotene hydroxylase and zeaxanthin glycosylase.
Here, GGPP is provided to the transformed host cells
via the nutrient medium as above, and the transformed
host cells convert the GGPP to the necessary phytoene
25   and then to lycopene, beta-carotene, zeaxanthin and
zeaxanthin diglucoside using the transformed structural
genes. Of course, cells are transformed with fewer
than all of the five genes where a carotenoid other
than zeaxanthin diglucoside is desired and GGPP is
30   provided by the medium to the cells.

It is understood in any of the methods of
carotenoid production contemplated herein that the
transformed host is free of exogenously supplied DNA
segments that inhibit the production and/or
35   accumulation of a desired carotenoid. Thus, for

-65-

example, where zeaxanthin is to be prepared,
exogenously supplied DNA that encodes a biologically
active zeaxanthin glycosylase that converts zeaxanthin
to zeaxanthin diglucoside is absent so that zeaxanthin
5   accumulates.

E. **Examples**
            The following examples are intended to
illustrate, but not limit, the scope of the invention.
10 Studies related to carotenoid biosynthesis generally,
GGPP synthase and phytoene synthase are discussed in
Examples 1-7, studies related to lycopene are discussed
in Examples 8-14, studies related to beta-carotene are
discussed in Examples 15-20, studies related to
15 zeaxanthin are discussed in Examples 21-25, whereas
Examples 26-30 discuss zeaxanthin diglucoside.
            All recombinant DNA techniques were performed
according to standard protocols as described in
Maniatis et al., _Molecular Cloning, A Laboratory_
20 _Manual_, Cold Spring Harbor Laboratory, Cold Spring
Harbor, NY (1982), except where noted. All restriction
enzymes and other enzymes were used according to the
supplier's instructions. DNA sequencing was performed
on M13 single-stranded DNA using a modification of the
25 basic dideoxy method of Sanger et al, _Proc. Natl. Acad._
Sci. U.S.A. 74:5463-7 (1977). A sequencing kit from
BRL Life Technologies, Inc., Gaithersburg, MD was used.
The DNA sequence was analyzed on the IG Suite from
Intelligenetics Corp.
30   Enzyme assays for enzymes engineered in
E. coli or _Saccharomyces cerevisiae_ were performed
according to the protocols provided in Example 2e for
GGPP synthase and phytoene synthase, in Example 8g for
phytoene dehydrogenase-4H, and in Example 15f for
35 lycopene cyclase.

-66-

Carotenoids were extracted and analyzed by high performance liquid chromatography (HPLC) from both E. coli or S. cerevisiae according to the protocol provided in Example 4. The identity of zeaxanthin diglucoside was confirmed by mass spectroscopy performed according to the protocol provided in Example 4. The identity of zeaxanthin was confirmed by mass spectroscopy. The identification of the other carotenoids was confirmed by elution from HPLC, UV-Visible spectral analysis, and comparison with known standards of phytoene, lycopene, and beta-carotene.

The method for production in E. coli of the proteins in E. coli encoded by the different genes, using the inducible Rec 7 promoter system in the plasmid pARC306A, is described in Example 2d. These proteins were used in the enzyme assays described. This protocol was also used to produce sufficient amounts of the proteins from which the N-terminus of the protein was determined.

Examples 21-25 discuss the production of zeaxanthin. Example 21 describes construction of an engineered, readily movable structural gene for beta-carotene hydroxylase, whereas Example 22 illustrates the incorporation of that structural gene into plasmid pARC306A to form plasmid pARC404BH, which when placed into E. coli cells along with plasmid pARC279 caused production of zeaxanthin.

Examples 26-30 discuss the production of zeaxanthin diglucoside. Example 26 describes construction of an engineered, readily movable structural gene for zeaxanthin glycosylase, which when placed into E. coli along with plasmid pARC288 caused production of zeaxanthin diglucoside.

-67-

<u>Example 1.</u>   Confirmation of the pr senc of the
                  carot noid biosynth sis pathway g n s in
                  <u>Erwinia</u> <u>herbicola</u> plasmid pARC376

      a. <u>E. coli</u>

      <u>E. coli</u> cells, which by themselves are not
capable of pigment formation, become intensely yellow
in color when transformed with plasmid pARC376 (Figure
5). The pigments responsible for the observed yellow
color were extracted from the cells and shown to be
zeaxanthin and zeaxanthin diglucosides from UV-VIS
spectral and mass spectral data.

      In the presence of diphenylamine in the growth
medium, pigment formation is strongly inhibited
resulting in colorless cells, which have been found to
accumulate trace amounts of phytoene. Diphenylamine is
known to inhibit the phytoene dehydrogenase-4H
reaction. This was the first indication that the
carotenoid pathway is functional in these transformed
cells. Harvesting mid-log phase cells and extracting
carotenoids from those cells indicated the presence of
phytoene, phytofluene, and zeta-carotene, further
confirming the presence of functional carotenoid
pathway syntheses in the cells.

      b. <u>A. tumefaciens</u>

      Carotenoid production in <u>A. tumefaciens</u>
containing the <u>Erwinia</u> <u>herbicola</u> carotenoid DNA was
investigated. Three plasmids containing various
portions of plasmid pARC376 were transformed into <u>A.</u>
<u>tumefaciens</u> strain LBA4404. Four different carotenoids
were produced, i.e., phytoene, lycopene, beta-carotene,
and zeaxanthin.

      The three plasmids used in this study were:
1. Plasmid pARC803 (about 17 kb), which contained the
    R1162 ori, the kanamycin resistance gene (NPTII)

-68-

and the _Erwinia herbicola_ DNA of plasmid pARC376-
Ava 103 fragment (derived by deleting 2 Ava I
restriction fragments, at about 8331-8842-10453,
and cloning the Hind III (about 13463) to Eco RI

5   (about 3370 Figure 5) fragment into plasmid pSOC925
(Figure 12);

2.   Plasmid pARC274 (about 17 kb), which contained the
R1162 ori, the kanamycin resistance gene, and the

10   _Erwinia herbicola_ DNA of plasmid pARC376-Bam 100
fragment (derived by deleting 2 Bam HI restriction
fragments, at about 3442-4487-5302 and cloning the
Hind III (about 13463) to Eco RI (about 3370,
Figure 5) fragment into plasmid pSOC925;

15

3.   Plasmid pARC288 (about 18 kb) which contained the
R1162 ori, the kanamycin resistance gene, the
Erwinia herbicola DNA of plasmid pARC376-Sal 8
(Example 2a) and the GGPP synthase gene fragment

20   from Hind III (about 13463) to Eco RV (about 11196,
Figure 5).

         These plasmids were transformed into competent
cells of _Agrobacterium_ according to the protocol below.

25   1.   An _Agrobacterium_ colony was grown overnight (about
15 hours) in 2 to 3 ml YP medium (10 g/l
Bactopeptone, 10 g/l yeast extracts, and 5 g/l
NaCl, pH 7).

2.   The overnight culture was transferred into 50 ml

30   fresh YP medium in 250 ml flask at 250 rpm and
28°C, and grown until the culture reached 0.5 to
1.0 OD (A_{600}).

3.   The culture was chilled on ice for 5 minutes, then
the cells were harvested by centrifugation.

-69-

4. The cells were resuspended in 1 ml of 20 mM calcium chloride.

5. About 1 µg of plasmid DNA was added into 0.1 ml of the cell suspension and mixture was incubated on

5   ice for 30 minutes.

6. The reaction mixture was frozen in liquid nitrogen for 1 to 2 minutes and then incubated at 37°C for 5 minutes.

7. One ml of YP medium was added and the mixture was

10   incubated at 28°C for 2 to 4 hours.

8. The cells were plated in LB medium (5 g/l yeast extracts, 10 g/l tryptone, 5 g/l NaCl, and 2 g/l glucose, pH 7) containing 50 µg/ml kanamycin.

The transformed cells were selected on LB

15   plates containing 50 µg/ml of kanamycin at 28°C (LB plates = 10 g/l tryptone, 5 g/l yeast extracts, 5 g/l NaCl, 2 g/l glucose, and 15 g/l Bactoagar). The transformed cells were cultivated on the same rich medium for two days, harvested and dried for carotenoid

20   extraction. For carotenoid extraction, 0.5 ml of water, 2.5 ml of acetone, and 2.5 ml of methanol were added to the dried cells. After 1 hour incubation with mixing at room temperature, the solvent containing carotenoids was filtered, and carotenoids isolated

25   were analyzed by HPLC.

The carotenoids produced by both E. coli and Agrobacterium are listed in Table 1. The amounts of carotenoids produced by Agrobacterium were about 5 to 10 times lower than by E. coli cells carrying the same

30   plasmids (by gross estimation).

-70-

Table 1
Carotenoids Produced by <u>A. tumefaciens</u> LBA4404

Plasmids	Major Carotenoids	
	E. coli	Agrobacterium
pARC803	Lycopene	Lycopene, Phytoene
pARC274	β-Carotene	β-Carotene, (Phytoene)*
pARC288	Zeaxanthin	Zeaxanthin

*   Minor component.

        The origin of replication from plasmid R1162,
described by Meyer, R. et al., <u>J. Bacteriol.</u> 152:140
(1982), was introduced into plasmid pARC376, to
construct a broad host-range plasmid capable of
replication in other bacteria. The resulting plasmid
was used to introduce <u>Erwinia herbicola</u> carotenoid DNA
into <u>Rhodobacter sphaeroides</u> and its carotenoid
mutants. The results demonstrated that the <u>Erwinia
herbicola</u> carotenoid DNA was not expressed in
<u>Rhodobacter</u> cells, presumably because there was no
complementation of the <u>Rhodobacter</u> phytoene synthase,
phytoene dehydrogenase-4H and neurosporene
dehydrogenase mutants. A further study, described
hereinafter, indicated that phytoene dehydrogenase-4H
could be expressed in <u>Rhodobacter</u> cells as hosts.

<u>Example 2.</u>   GGPP Synthase Gene
        The GGPP synthase gene was obtained from the
pARC376 plasmid utilizing the following methods.

-71-

a. Digestion of pARC376 with Sal I

The plasmid pARC376-Sal 8 is a derivative of plasmid pARC376 from which two Sal I fragments were removed. One of those fragments is the approximately
5   1092 bp fragment bounded by the Sal I restriction sites at about 9340 and about 10432 shown in Figure 5, whereas the other is the 3831 bp (approximate size) fragment bounded by the Sal I restriction sites at about 10432 and about 14263 also in Figure 5. This was
10 accomplished as follows.

Plasmid pARC376 DNA was prepared using the alkaline lysis method. 5 Micrograms of plasmid DNA were digested with Sal I (BRL) in a high salt buffer provided by the supplier and additionally containing
15 150 mM NaCl, for 1 hour at 37°C and purified on a 0.8 percent agarose gel. The remaining plasmid, about 10.2 kilobases in length, was electroeluted from the gel, phenol extracted and ethanol precipitated. After elimination of the above Sal I fragments from about
20 positions 9340 to 14263, the remaining DNA was religated to itself to form plasmid pARC376-Sal 8.

b. Construction of pARC808

To determine if the gene for GGPP synthase was
25 present on the deleted Erwinia herbicola DNA, plasmid pARC376-Sal 8 was cloned into plasmid pSOC925, an E. coli plasmid R1162 derivative, to generate plasmid pARC808. The plasmid pSOC925 contains the origin of replication from the R1162 plasmid, the NPT II gene
30 from Tn5 that confers resistance to kanamycin, and unique Hind III and Eco RI restriction sites.

Briefly, the plasmid pSOC925 expression DNA vector was prepared for cloning by admixing 5 µg of plasmid DNA to a solution containing 5 units of each of
35 the restriction endonucleases Hind III and Eco RI and

-72-

the Medium Salt Buffer from Maniatis. This solution
was maintained at 37°C for 2 hours. The solution was
heated at 65°C to inactivate the restriction
endonucleases. The DNA was purified by extracting the
5   solution with a mixture of phenol and chloroform
followed by ethanol precipitation.

Plasmid pARC376-Sal 8 was digested with Hind
III and Eco RI in a similar way. The Erwinia herbicola
DNA in plasmid pARC376-Sal 8 from the Hind III site at
10   about position 348 to the Eco RI site at about position
3370 (Figure 5) was then ligated into the plasmid
vector pSOC925 that had already been digested with Hind
III and Eco RI.

The ligation reaction contained about 0.1 µg
15   of the plasmid vector pSOC925 and about 0.2 µg of the
Erwinia herbicola Hind III to Eco RI fragment from
plasmid pARC376-Sal 8 in a volume of 18 µl. Two µl of
10 X ligation buffer (IBI, Corp) and 2 units of T4
ligase were added. The ligation reaction was incubated
20   at 4°C overnight (about 15 hours). The ligated DNA was
transformed into E. coli HB 101 according to standard
procedures (Maniatis). This generated the plasmid
pARC808, which also codes for kanamycin resistance.
The excised DNA fragment from plasmid pARC376-Sal 8
25   contains an endogenous promoter sequence upstream from
the GGPP synthase gene.

Positive clones with inserts were identified
by growing prospective positive clones, isolating
plasmid DNA by the alkali lysis method (Maniatis), and
30   performing restriction enzyme analysis on the isolated
plasmid DNA's. E. coli cells transformed with this
plasmid DNA did not produce colored carotenoids, as
determined by visual inspection and HPLC and TLC
analysis. Other studies discussed hereinafter
35   demonstrated that plasmid pARC808 expresses Erwinia

<u>herbicola</u> enzymes that can convert phytoene into
colored carotenoid pigments.

c. Construction of pARC282

A second plasmid was constructed by inserting
a restriction fragment containing the approximately
1153 bp Bgl II (about position 12349, Figure 5) to Eco
RV (about position 11196, Figure 5) fragment from
plasmid pARC376 into the Bam HI and Hind III sites of
pBR322 to produce plasmid pARC282. Briefly, the
plasmid pARC273 contains the <u>Erwinia</u> <u>herbicola</u> DNA from
the Bgl II site (at about position 12349) to the Eco RV
site (at about position 11196).

About 100 non-coding bp downstream from the
Eco RV site in plasmid pARC273 is a Hind III
restriction site, which is a part of the pARC273
vector. Here, about 5 μg of the plasmid pARC273 were
incubated with 5 units of each of the restriction
enzymes Bgl II and Hind III in the Medium Salt Buffer
(Maniatis) for 2 hours at 37°C. Five μg of the vector
pBR322 were incubated with 5 units of each of the
restriction enzymes Bam HI and Hind III in the Medium
Salt Buffer (Maniatis) for 2 hours at 37°C.

The <u>Erwinia</u> <u>herbicola</u> Bgl II to Hind III DNA
fragment (about 0.2 mg) from plasmid pARC273 was
admixed with the Bam HI and Hind III digested plasmid
pBR322 vector (about 0.1 μg) in 18 μl total volume.
Two μl of 10 X Ligation Buffer (IBI, Corp.) and 2 units
of T$_4$ Ligase were added, the reaction was incubated
overnight (about 15 hours) at 4°C, and the ligated DNA
was transformed into competent <u>E. coli</u> HB101 cells
according to procedures in Maniatis. Positive clones
were identified by growing the prospective
transformants, isolating plasmid DNA by the alkali

-74-

lysis method (Maniatis), and performing restriction
enzyme analysis on the plasmid DNA.

This plasmid, pARC282 encodes ampicillin
resistance in E. coli and includes a native Erwinia
5   herbicola promoter between the Bgl II site and the
initial Met codon of the GGPP synthase gene, but does
not cause any carotenoids to be produced. However,
when this plasmid was transferred into E. coli cells
containing the plasmid pARC808, and the E. coli cells
10   were grown in the presence of both kanamycin and
ampicillin, carotenoids were synthesized as evidenced
by production of the yellow pigment zeaxanthin. Thus,
plasmid pARC282 contained the essential gene that was
deleted from the plasmid pARC376-Sal 8 plasmid, and the
15   presence of this gene in combination with other Erwinia
herbicola carotenoid genes could restore carotenoid
production in E. coli.

d.   Other Plasmid Constructs

20   Enzyme assays were performed on similar
plasmid constructs, including plasmid pARC491 which was
constructed by cloning the approximately 1068 bp
fragment from Hpa I (at about position 12264 of plasmid
pARC376 or about position 84 of Figure 2) to Eco RV (at
25   about position 11196, Figure 5) into a plasmid
denominated pARC306A. Plasmid pARC306A, whose
restriction map is illustrated in Figure 6 contains
approximately 2519 base pairs. This plasmid contains
the polylinker region from pUC18, a unique Nco I site,
30   the ampicillin selectable marker, the pMB1 origin of
replication and the Rec 7 promoter. Cells containing
this plasmid construct had a level of 7.91 nmol/min/mg
protein activity of GGPP synthase.

35

-75-

. DNA sequ ncing

The accuracy of some of the cloning steps was
confirmed by sequencing the insert using the dideoxy
method described by Sanger et al., <u>Proc. Natl. Acad.</u>
<u>Sci</u>. USA, 74:5463-5467, (1977) and following the
manufacturer's instructions included in a sequencing
kit from BRL.

The DNA sequence was determined for the
approximately 1153 base pair restriction fragment from
the region between the Bgl II site at about 12349 of
Figure 5 and the Eco RV site at about 11196 of Figure
5. The obtained DNA sequence and putative partial
amino acid residue sequences are shown in Figure 2
(about positions 1 to 1153). The direction of
transcription of the gene for GGPP synthase in plasmid
pARC376 (Figure 5) is counterclockwise and proceeds in
the direction from the Bgl II site toward the Eco RV
site.

f.   In Vitro mutagenesis

The initiation codon for GGPP synthase begins
at about nucleotide position 12226 of plasmid pARC376
with the ATG codon for methionine (about position 124
of Figure 2). A Nco I restriction site was introduced
at this position of the GGPP synthase gene using <u>in</u>
<u>vitro</u> mutagenesis following the techniques described in
<u>Current Protocols In Molecular Biology</u>, Ausabel et al.
eds., John Wiley & Sons, New York, (1987) p. 8.1.1-
8.1.6, with the exception that <u>E. coli</u> CJ 236 was grown
(in step 3 at page 8.1.1) in further presence of 20
μg/μl chloramphenicol. The primer used was:

5'   3'
TCA GCG GGT AAC CTT G<u>CC ATG G</u>GG AGT GGC AGT AAA GCG
                      Nco I site   (SEQ ID NO:18)

-76-

The mutations were confirmed either by DNA
sequencing or by the presence of the newly introduced
Nco I site. This manipulation changed the natural
sequence

5   TTG CAATGG TGA   (SEQ ID NO:19) to

               TTG CCATGG GGA,   (SEQ ID NO:20)

wherein a bold-faced letter above and in the following
examples indicates an altered base.

This modified version of the GGPP synthase
10   gene from the newly introduced Nco I site to the Eco RV
site (about 1029 bp) was then inserted into the plasmid
pARC306A to generate plasmid pARC417BH. This plasmid,
pARC417BH, contains the E. coli promoter Rec 7 adjacent
to a multiple cloning site. Structural genes lacking a
15   promoter region, when introduced adjacent to the Rec 7
promoter, are expressed in E. coli.

When plasmid pARC417BH was introduced into E.
coli cells, GGPP synthase enzyme activity (measured as
GGOH) was found at the level of 6.35 nmol/min/mg
20   protein. In addition, when plasmid pARC417BH was
introduced into E. coli cells containing plasmid
pARC808, carotenoids were produced. This demonstrated
that the gene for GGPP synthase had been identified and
genetically engineered.

25

          g.   Fine tuning the GGPP synthase gene
               Several constructs designed to express the
GGPP synthase gene were made to optimize the expression
of an active GGPP synthase enzyme. Again using in
30   vitro mutagenesis according to methods previously
cited, a Nco I site was introduced at about position
12264 of plasmid pARC376, using the primer,

      5'   3'
35   CAT GGC GAA ATA GAA GCC ATG GGA CAA TCC ATT GAC GAT
                         Nco I site   (SEQ ID NO:21)

-77-

17 amino acids downstream from the initiation codon for
the GGPP synthase gene that is located at about
position 124 in Figure 2. That site was thus placed at
the upstream side of the MET whose ATG codon begins at
about position 175 of the sequence of Figure 2. The
natural DNA sequence

              AAG TAATGA GAC   (SEQ ID NO:22) was changed to

              AAG **CCATGG** GAC.   (SEQ ID NO:23)

This modified GGPP synthase gene coding for seventeen
fewer amino-terminal amino acid residues was inserted
into plasmid pARC306A at the Nco I site of that plasmid
to generate plasmid pARC418BH.

     When GGPP synthase assays were performed on
cells transformed with plasmid pARC418BH, no enzyme
activity was detected. In addition, when this modified
GGPP synthase was added to E. coli cells containing the
plasmid having the rest of the genes for the enzymes
required for carotenoid synthesis, plasmid pARC808
described above, no carotenoids were synthesized. This
demonstrated that deletion of the 17 N-terminal amino
acids of the GGPP synthase resulted in a non-functional
enzyme.

     A final construction was made at the 5' end of
the GGPP synthase gene. A fragment excised from the
GGPP synthase gene from the Nru I site (about 12187 of
plasmid pARC376) through the Eco RV site (about 11196
of plasmid pARC376 or at about positions 162 through
1153 in Figure 2) to the Hind III site of plasmid
pARC282 was inserted into the pARC306A plasmid, to form
plasmid pARC489B that is discussed below.

     Plasmid pARC306A was digested with Eco RI.
The Eco RI end was converted to a blunt end using the
Klenow fragment of DNA Pol I according to the usual
techniques described by Maniatis. The Nru I blunt 5'
end from the partially Nru I-Hind III fragment of

-78-

plasmid pARC282 was blunt end-ligated to the newly
generated blunt end at the Eco RI site of plasmid
pARC306A by mixing about 0.2 µg of the Nru I to Eco RV
Erwinia herbicola DNA with the pARC306A vector in about

5   18 µl volume. Two µl of 10X Ligation Buffer (IBI,
Corp.) and two units of T4 Ligase were added, and the
reaction was incubated at 4°C overnight (about 15
hours).

The resulting partially ligated plasmid was

10   then digested with Hind III, which resulted in the loss
of the polylinker region shown in Figure 6 from the Eco
RI site to the Hind III site. The resulting Hind III
sticky ends were then ligated to form plasmid pARC489B.

Positive clones were identified by plasmid DNA

15   isolation (Maniatis), and by restriction enzyme
analysis on the plasmid DNA.

In plasmid pARC489B, DNA coding for the first
13 amino acid residues of the GGPP gene was deleted.
The first four amino acid residues encoded downstream

20   from the Rec 7 promoter in plasmid pARC306A and the
newly generated Eco RI blunt end were placed upstream
from the former Nru I site of GGPP synthase. This
altered the N-terminal amino acid sequence of GGPP
synthase in the following manner. The difference in

25   amino acid sequence became:
Original Amino Acid Sequence of Native Erwinia
herbicola GGPP Synthase.

MET VAL SER GLY SER LYS ALA GLY VAL SER PRO HIS ARG

GLU ILE...

30   (SEQ ID NO:24)
Amino Acid Sequence of modified GGPP Synthase Gene in
Plasmid pARC489B

MET ALA GLU PHE GLU ILE...

(SEQ ID NO:25)

The determined DNA sequence for this het rologous gene
is illustrated in Figure 3 from about position 150 to
about position 1153.

5   E. coli cells transformed with the plasmid
pARC489B were assayed for GGPP synthase activity. The
level of activity was found to be 12.15 nmol/min/mg
protein.

        When the plasmid pARC489B was transferred to
E. coli cells that contained a plasmid containing the

10   rest of the genes coding for enzymes required for
carotenoid production, plasmid pARC808, the cells
produced carotenoids. Therefore, this construction
coded for an active enzyme even though the heterologous
gene portion from plasmid pARC306A encoded the first

15   four amino acid residues, and the first 13 amino acid
residues encoded by the gene for GGPP synthase were
deleted.

        The above described DNA segment of plasmid
pARC489B overlaps bases encoding four amino acids

20   adjacent to the Rec 7 promoter at its 5' end and
extends to the blunted, former Eco RI site in the
polylinker region of the plasmid. This DNA segment can
be excised by reaction with Nco I at its 5' end and the
Hind III or Pvu II sites as are illustrated for plasmid

25   pARC306A in Figure 6.

        The desired GGPP synthase gene does not
contain a Pvu II or a Hind III restriction site. The
region between the Hind III and Pvu II sites of plasmid
pARC489B contains stop codons in all three reading

30   frames. It is preferred to utilize the Pvu II site for
cleavage of the 3' end of the DNA. Thus, the desired
GGPP synthase DNA segment can be referred to as lying
within the approximately 1150 bp sequence between the
Nco I and Pvu II restriction sites of plasmid pARC489B.

-80-

Next, the 3' end of the gene for GGPP synthase
was modified. This construction was made in the
following manner. A Nru I (about 11187)-Bal I (about
11347 of Figure 5) double blunt end fragment was
5   inserted into a specially prepared version of the
plasmid pARC306A. Thus, plasmid pARC306A was digested
with Eco RI, and the resulting ends were blunted with
the Klenow fragment of DNA Pol I to produce two blunt
ends. The blunt ended Nru I-Bal I fragment was
10   operatively linked into the plasmid pARC306A vector to
form plasmid pARC489D. This vector included all of the
polylinker restriction sites of plasmid pARC306A shown
in Figure 6 except the Eco RI site.

        The GGPP synthase gene-containing portion of
15   the resulting plasmid pARC489D has the same 5' end as
does plasmid pARC489B, but the 3' end is about 151 bp
shorter than the GGPP synthase gene in plasmid
pARC489B. The sequence of the heterologous GGPP
synthase structural gene of plasmid pARC489D is
20   illustrated in Figure 3 from about position 150 to
about position 1002, with the 5' end of this DNA being
the same as that of the GGPP synthase gene present in
plasmid pARC489B.

        Downstream about 70 bp from the Hind III site
25   of the multiple cloning region in plasmid pARC306A is a
Pvu II site. There are no Pvu II sites in the GGPP
synthase gene. Therefore, the GGPP synthase structural
gene can be transferred from a pARC306A-derived plasmid
such as plasmid pARC489D to other plasmids as an
30   approximately 1000 bp Nco I-Pvu II fragment.

        Plasmid pARC489D was transformed into E. coli.
Very surprisingly, this construction gave the highest
enzyme activity of all the different versions of the
GGPP synthase gene. This activity was an unexpectedly
35   high 23.28 nmol/min/mg protein.

When the plasmid pARC489D was introduced into
E. coli cells containing the plasmid pARC808,
carotenoids were synthesized.

A comparison of the activities of several of
the previously described GGPP synthase gene constructs
is shown in Table 2 below, including the activity of a
related gene present inherently in R. sphaeroides
2.4.1. Those results indicate an enhancement of about
35 to about 130 times the activity of the original
plasmid pARC376.

Table 2

GGPP Synthase Activity of Various Gene Constructs
As Compared to R. sphaeroides

Constructs	Activity (nmol/min/mg protein)
R. sphaeroides 2.4.1	0.20
pARC376	0.18
pARC491	7.91
pARC417BH	6.35
pARC418BH	0
pARC489B	12.15
pARC489D	23.28

h. GGPP synthase characterization

The plasmids pARC489B and pARC489D were
introduced into the E. coli strain JM101 (BRL). These
cells were treated with nalidixic acid to induce the
Rec 7 promoter, which caused production of large
amounts of the GGPP synthase enzyme. The protein
extract from these cells was separated on SDS-
polyacrylamide gel electrophoresis (PAGE). Because of
the very large amount of GGPP synthase produced under

-82-

these conditions, it is readily identifiable by
stain.g with Coomassie Brilliant Blue on the SDS-PAGE
system. The isolated and substantially purified GGPP
synthase can then be recovered from the gels by
5   standard procedures.

The *Erwinia* *herbicola* GGPP synthase that was
produced in cells containing plasmid pARC489B was a
protein of the size of about 35 kilodaltons, and is
thought to be the complete, native molecule, whereas
10   the GGPP synthase that was produced in cells with
plasmid pARC489D was about 33 kilodaltons. Thus, the
5' deletion of thirteen amino acid residues and then
replacement with non-*Erwinia* *herbicola* sequence of four
residues, coupled with the 3' deletion of the
15   approximately 151 bp between the Bal I site and the Eco
RV site produced a protein that was about 2 kilodaltons
smaller, but far more active than the native molecule.
The GGPP synthase structural gene present in plasmid
pARC489D is the gene most preferably used for GGPP
20   synthase in *E.* *coli*, *S.* *cerevisiae*, and higher plants.

       i.   **Induction of Rec 7 driven protein
            production**

The previously discussed production of GGPP
25   synthase in *E.* *coli* using plasmids pARC417BH, pARC489B
and pARC489D was carried out using the Rec 7 promoter.
Phytoene synthase production in *E.* *coli* using the
plasmid pARC140N discussed below was also carried out
using the Rec 7 promoter. Culture conditions for
30   growth of the transformed *E.* *coli* cells are as follows.

A single colony from a plate containing
freshly (<2 days old) transformed cells was picked,
grown overnight (e.g. about 15-18 hours) in M9+CAGM
medium (see Table 3B hereinafter for media
35   formulations) + 50 μg/ml ampicillin at 30°C. Cultures

-83-

of cells were grown at various temperatures from
27-37°C by diluting the cells 1:100 into fresh M9+CAGM
medium and maintaining the culture at the desired
temperature. Each culture was grown until it was
5   roughly one-half of the final desired density (150-180
Klett units in a shaken culture). The culture was then
induced by addition of nalidixic acid to a final
concentration of 50 μg/ml. Five μl of a stock solution
of freshly prepared 10 mg/ml nalidixic acid in 0.1N
10   NaOH per ml of culture to be induced was used.
Induction was permitted to proceed for 2-4 hours after
addition of nalidixic acid.

−84−

Table 3

A.   M9+CAGM MEDIUM COMPOSITION

Component	grams/liter
$Na_2HPO4.7H_{2O}$	13.2
KH_2PO4	3.0
NaCl	0.5
NH_4Cl	1.0
Casamino Acids (Difco)	10.0
$MgSO_4$	0.3
$CaCl_2.2H_2O$	0.004
Glucose (Shake Flask)	3.0
Thiamine-HCl	0.025
$FeCl_3$	0.0054
$ZnSO_4$	0.0004
$CoCl_2$	0.0007
Na_2MoO_4	0.0007
$CuSO_4$	0.0008
H_2BO_3	0.0002
$MnSO_4$	0.0005

B.   MEDIUM FORMULATIONS

M9+CAGM Medium for Shake Flasks   (1 Liter)

900 ml	distilled H_2	Autoclaved
40 ml	25X M9 Salts	Autoclaved
50 ml	20% (w/v) Casamino Acids	Filtered
6.4 ml	40% (w/v) Glucose	Autoclaved
1.2 ml	1M MgSO	Autoclaved
0.25 ml	0.1M $CaCl_2$	Autoclaved
0.25 ml	0.1% (w/v) Thiamine-HCl	Filtered
0.1 ml	10,000X Trace Minerals	Filtered
0.1 ml	10,000X Iron Supplement	Filtered

-85-

All components should be sterilized separately, cooled to room temperature and then combined.

C. <u>25X M9 Salts   (1 liter)</u>

Component	grams
Na2HPO4.7H$_2$O	330
KH$_2$PO$_4$	75
NH$_{4}$Cl	25

distilled H$_2$O to 1 Liter

D. <u>10,000X Trace Minerals   (200 ml)</u>

Component	grams
ZnSO$_4$	0.8
COCl$_2$	1.4
Na$_2$MoO$_4$	1.4
CuSO$_4$	1.6
H$_2$BO$_3$	0.4
MnSO$_4$	1.0

Dissolve in 200 ml of H$_2$O, add 1 drop HCl (fuming), filter sterilize.

E. <u>10,000X Iron Supplement   (200 ml)</u>

Component	grams
FeCl$_3$	10.8

Dissolve in 200 ml of H$_2$O, add 1 drop HCl (fuming), filter sterilize.

Each culture was highly aerated at all times. Fifteen ml in a 250 ml sidearm flask for analytical runs were routinely used, and 330 ml in a Fernbach (2.81) flask for semi-preparative runs were routinely used.

Production of all proteins examined so far has
been quite dependent on strong aeration during
the induction period.

5   **j. Enzyme assay**
GGPP synthase was prepared in the cell cytosol
as described below.

(1) Cytosol preparation
10   The growing cells were centrifuged to form a
cell pellet. The cell pellet was resuspended in 50 mM
potassium phosphate buffer, pH 7.0, containing 10
percent glycerol, 0.1 mM EDTA in a 15 ml plastic
conical tube and vortexed with acid washed glass beads
15   (425-600 micron for yeast cells and 75-150 micron for
bacteria are typically used) for 1 minute and allowed
to cool in ice for 1 minute. This was repeated three
times after which the homogenate was transferred to
another tube and centrifuged at 17,000 x g for 60
20   minutes at 4°C. The supernatant was next centrifuged
at 150,000 x g for 60 minutes at 4°C. The supernatant
thus obtained was the cell cytosol.

(2) Assay for GGPP synthase
25   Cell cytosol was preincubated for 20 minutes
at 4°C with 10μM epoxy-isopentenyl pyrophosphate (IPP)
in order to inhibit IPP-isomerase activity. The assay
mixture, containing 40 μM farnesyl pyrophosphate (FPP)
and 40 μM 14C-IPP (250,000 dpm) in 10 mM Hepes buffer
30   (pH 7.0, 1 mM MgCl$_2$, 1 mM DTT) in a 1 ml total volume
of preincubated cytosol, was incubated at 37°C for 30
minutes.
The reaction was terminated by transferring
the assay mixture to a pre-cooled (in dry ice) tube and
35   lyophilizing for 8 hours. The dry residue was

resuspended in 0.5 ml of 0.1 M glycine buffer (pH 10.4,
1 mM $MgCl_2$, 1 mM $ZnCl_2$) and treated with 25 units of
alkaline phosphatase for 3 hours at 37°C. The alkaline
phosphatase reaction converted the pyrophosphates to
their corresponding alcohols, which were extracted with
hexane, evaporated to dryness under a stream of
nitrogen and redissolved in 150 μl of methanol.

Seventy-five μl of this methanol solution were
injected into an HPLC connected with a C-18 econosphere
Altech analytical column (4.6 x 250 mm, 5 micron
particle size) equilibrated with 85 percent
methanol:water (4:1) and 15 percent $THF:CH_3CN$ (1:1). A
linear gradient to 80 percent methanol:water (4:1) and
20 percent $THF:CH_{3C}N$ (1:1) in 20 minutes at 1.5 ml/min
resolved the alcohols. The HPLC was connected in
series with a Radiomatic flow detector, which
integrated the radioactive peaks, e.g. geranylgeraniol
(GGOH) peak. Specific activity was expressed in nmol
GGOH formed/min/mg of protein under the given assay
conditions. Protein was determined by the Bradford
method using BSA as the standard.

Example 3. Phytoene Synthase Gene
a. Digestion of pARC376 with Pst I

The plasmid pARC376-Pst 122 was created by
deletion of an approximately 592 bp Pst I <u>Erwinia
herbicola</u> DNA fragment from Pst I sites at about 5807
to about 5215 of plasmid pARC376 (Figure 5), followed
by religation of the larger of the two fragments. The
Eco RI (about 3370) to Hind III (about 13463) fragment
from plasmid pARC376-Pst 122, which contains the
desired <u>Erwinia herbicola</u> DNA fragment, was cloned into
the plasmid pARC305A, resulting in plasmid pARC139.

The plasmid pARC305A contains the polycloning
linker from pUC18, the chloramphenicol

acetyltransferase gene (CAT) that confers
chloramphenicol resistance in E. coli and the pMB1
origin of replication. The plasmid pARC305A is an
analogous plasmid to plasmid pUC18 except plasmid

5   pARC305A contains the CAT selectable marker whereas
pUC18 contains the ampicillin selectable marker.
       When the resulting Erwinia herbicola DNA was
inserted into the plasmid pARC305A to create the
plasmid pARC139 and introduced into E. coli cells, no

10   carotenoids were made, as expected.
       An impairment of the gene for phytoene
synthase would cause the E. coli cells not to produce
any colored carotenoids. Therefore, the deletion of
this 592 bp region could have deleted part of the gene

15   for phytoene synthase.

       b.   Construction of Plasmid pARC285
       The construction of plasmid pARC285 used the
approximately 1112 bp Nco I to Eco RI fragment from the

20   plasmid pARC376-Bam 100. The plasmid pARC376-Bam 100
is a derivative of the pARC376 plasmid in which the
approximately 1045 bp Bam HI fragment from about
position 3442 to about position 4482 (Figure 5) and the
approximately 815 bp Bam HI fragment from about

25   position 4487 to about 5302 (Figure 5) were deleted. A
total of about 1860 nucleotides was deleted from the
pARC376 plasmid. As a result of the deletions of the
Bam HI fragments from plasmid pARC376, the Bam HI site
at about 5354 at the 3' end was brought within about 72

30   nucleotides of the Eco RI site originally at about
position 3370 of plasmid pARC376. The resulting
restriction fragment therefore contained about 1112 bp
and was bounded by Nco I and Eco RI restriction sites
at its 5' and 3' ends, respectively.

-89-

The phytoene synthase gene is contained on an approximately 1040 bp Nco I to Bam HI restriction fragment (corresponding approximately to positions 6342 and 5302 of Figure 5, respectively), but it can be cloned into other plasmids as an approximately 1112 bp Nco I to Eco RI fragment. The approximately 1112 bp Nco I to Eco RI fragment was excised from the plasmid pARC376-Bam 100 and cloned into the Nco I to Eco RI sites of plasmid pARC306A to generate plasmid pARC285. The relevant portion of the phytoene synthase gene can thus be excised from plasmid pARC285 as an approximately 1112 bp Nco I to Eco RI fragment.

c.   Construction of Plasmid pARC140N

Analysis of the region surrounding the Nco I (about position 6342) site revealed that the methionine codon internal to the Nco I site was in an open reading frame that had another methionine codon 13 amino acid residues upstream. Immediately upstream from this methionine codon, was a consensus sequence for the ribosome binding site (AGGA) that is often found in procaryotic organisms upstream from the initiation codon of a gene.

To determine if the upstream methionine was in fact the initiation codon, a Bgl II site was introduced immediately upstream from the methionine codon of the Nco I site, using _in vitro_ mutagenesis, as described before. Two complementary polynucleotide sequences were made that contained a Nco I overhang on one end and on the other end a Bgl II overhang. The sequences were as follows:

Bgl II   Nco I
    5'   (SEQ ID NO:26)   3'
    GATCTAAAATGAGCCAACCGCCGCTGCTTGACCACGCCACGCAGAC
              ATTTTACTCGGTTGGCGGCGACGAACTGGTGCGGTGCGTCTGGTAC
5   3'   (SEQ ID NO:27)   5'

The two complementary single stranded polynucleotide sequences were hybridized together, ligated to an approximately 1112 bp Nco I-Eco RI fragment from plasmid pARC285 containing the

10   approximately 1040 bp Nco I to Bam HI phytoene synthase gene region and cloned into plasmid pARC135.

The plasmid pARC135 (shown in Figure 7) is composed of the pUC18 vector containing the yeast PGK promoter and terminator sequences separated by a unique

15   Bgl II site.

First, the approximately 3.1 kb Hind III fragment of yeast (S. cerevisiae) containing the PGK gene was cloned into the Hind III site of pUC18 to create plasmid pSOC117 (also referred to herein as

20   pARC117). Next, a Bgl II site was introduced by oligonucleotide mutagenesis upstream of the initiating ATG codon of the PGK gene contained within a mp19M13 clone, producing the change shown below in bold.

25   Native PGK Sequence: Met Ser Leu
          ....ACAACAAAATATAAAAACA ATG TCT TTA
                                        (SEQ ID NO:28)
      New PGK Sequence:
          ....ACAACAAGATCTAAAAACA ATG TCT TTA
30   (SEQ ID NO:29)
                Bgl II Site

Then, an approximately 1.1 kb Bst XI fragment, carrying

35   the introduced Bgl II PGK site, was excised from the mp19 clone and used to replace the homologous Bst XI fragment within plasmid pSOC117. Finally, the Bgl II fragment, containing the majority of the PGK structural

gene, was removed by Bgl II digestion, and the plasmid
was religated to yield plasmid pARC135. Plasmid
pARC135 was digested with Nco I and Eco RI, the
resulting gene was thereafter manipulated, as discussed
5   below, to generate the plasmid pARC140R, which contains
the <u>S. cerevisiae</u> phosphoglyceric acid kinase (<u>PGK</u>)
promoter at the Bgl II site.
        The experimental protocol for the construction
of plasmid pARC140R is described below.
10   A. Hybridization/Annealing of the two
                oligonucleotide probes (oligonucleotide probes
                were not phosphorylated at the 5' end).
                1) The two complementary oligonucleotide probes
                    were annealed in 25 μl of solution
15   containing:
                        10 μl of oligonucleotide #1 (about 1 μg)
                        10 μl of oligonucleotide #2 (about 1 μg)
                        1.65 μl of 1 M Tris-Cl$_2$ (pH 8.0)
                        2.5 μl of 100 mM MgCl$_2$
20   0.45 μl water
                2) The probe solution was incubated at 65°C for
                    10 minutes. Then it was cooled according to
                    the following regime:
                    20 minutes at 55°C
25   20 minutes at 42°C
                    20 minutes at 37°C
                    30 minutes at room temperature (24°C)
            B. An approximately 1112 bp fragment from Nco I to
                Eco RI in plasmid pARC285, containing an
30   approximately 1040 bp (Nco I to Bam HI) sequence
                was excised and isolated from the gel. This
                approximately 1112 bp fragment contained the
                shortened version of the gene for phytoene
                synthase.

-92-

C. The annealed oligonucleotide probes were ligated
    overnight (15 hours at 15°C) to the
    approximately 1113 bp (Nco I to Eco RI) fragment
    according to the following protocol:

5

                Annealed oligos   25 µl
                Nco I-Eco RI fragment   20 µl (about 1 µg)
                10 X Ligation Buffer   5 µl (IBI, Corp.)
                T4 Ligase (Boerhinger-Mannheim)

10 The result from the ligation was the following:

    Bgl II Nco I   Bam HI Eco RI
    I-------I-----------------------I-------I

15   D. The mixture was subsequently phenol extracted,
            chloroform: isoamyl alcohol (24:1) extracted and
            then ethanol precipitated. The DNA pellet was
            resuspended in 27 µl water.
        E. The DNA pellet was then digested for 30 minutes
20   at 37°C with Eco RI to remove any dimers that
            may have formed during the ligations.
                DNA fragment   27 µl
                Eco RI digestion buffer (BRL) 3 µl
                Eco RI enzyme (BRL)   3 µl (30 U)
25   F. The products of the Eco RI digestion were
            separated by electrophoresis on a 0.7 percent
            agarose gel. The fragment (about 1158 bp) was
            isolated from the gel.
        G. This Bgl II to Eco RI fragment was cloned into
30   the Bgl II and Eco RI sites of the plasmid
            pARC135 as follows. About 5 µg of plasmid
            pARC135 was digested with Bgl II and Eco RI and
            then separated on a 0.7 percent agarose gel. A
            DNA fragment (about 4 kb) was isolated. The
35   approximately 1158 bp Bgl II to Eco RI fragment

-93-

containing the full length phytoene synthase
gene was cloned into the approximately 4 kb
vector in the Bgl II and Eco RI sites according
to the following protocol:

5

pARC135 Bgl II/Eco RI digested   10 μl (about 0.2 μg)
Bgl II to Eco RI fragment   20 μl (about 0.5 μg)
10 X ligation buffer   3 μl
T4 ligase   2 μl (4 Units)

10   The reaction was incubated overnight (about 15-
18 hours) at 15°C.
        H. The ligated DNA was cloned into DH5-alpha
            E. coli cells obtained from BRL.
        I. Transformants were grown in the presence of 100
15   μg/ml of ampicillin. Colonies containing the
            cloned DNA fragment were identified by growing
            prospective clones in the presence of
            ampicillin, isolating plasmid DNA by the alkali
            lysis procedure and performing restriction
20   enzyme analysis on the clones. The result of
            this cloning procedure was a plasmid named
            pARC140R that contained the desired genes.

            Upstream from the ATG methionine codon, three
25   adenine residues were introduced. Presence of adenine
        residues adjacent to the initiation codon has been
        correlated with genes that are highly expressed in
        S. cerevisiae. These residues had been inserted in the
        sequence to cause high level expression of a gene in
30   S. cerevisiae (Hamilton et al., Nucleic Acids Research,
        15:3581 1987). The plasmid pARC140R contains the
        S. cerevisiae promoter from the gene for
        phosphoglyceric acid kinase (PGK) adjacent to the gene
        for phytoene synthase.

-94-

        The modified phytoene synthase structural gene
was excised from plasmid pARC140R as an approximately
1158 bp Bgl II-Eco RI fragment, engineered and cloned
into plasmid pARC306N to generate plasmid pARC140N.
5   The plasmid pARC306N is similar to plasmid pARC306A
except that instead of an Nco I site adjacent to the E.
coli Rec 7 promoter, there is an Nde I site.
        More specifically, plasmid pARC306N was
digested with Nde I and then digested with S1 nuclease
10   to blunt the ends of the former Nde I sites. The
plasmid was thereafter digested with Eco RI to remove
one of the blunt ends and provide an Eco RI sticky end.
        Plasmid pARC140R was digested with Bgl II and
then with S1 nuclease to blunt the resulting ends. The
15   digested and blunt-ended plasmid was then further
digested with Eco RI to remove one of the blunt ends
and provide an Eco RI sticky end for the DNA containing
the phytoene synthase structural gene. That structural
gene was therefore in a fragment of about 1164 bp with
20   a blunt end at one end and an Eco RI site at the other
end.
        The above phytoene synthase structural gene-
containing DNA segment was ligated into the blunt end
and to Eco RI portions of the above-digested plasmid
25   pARC306N to operatively link the two DNA segments
together and form plasmid pARC140N. The phytoene
synthase structural gene-containing DNA segment can be
excised from plasmid pARC140N as an approximately 1176
bp Hpa I-Eco RI fragment, an approximately 1238 bp Pvu
30   II-Eco RI fragment or as a still larger fragment using
one of the restriction sites in the polylinker region
downstream from the Eco RI site (see, Figure 6).
        The plasmid pARC140N, was transferred into
E. coli cells that contained the plasmid pARC139, in
35   which part of the gene for phytoene synthase was

deleted and, those E. coli cells were therefore
incapable of producing any colored carotenoids. When
plasmid pARC140N was added to those E. coli cells
containing plasmid pARC139, the cells produced colored

5   carotenoids. This demonstrated that the modified gene
for phytoene synthase coded for a functional enzyme.

E. coli cells containing plasmid pARC140N were
induced with nalidixic acid to produce large amounts of
the phytoene synthase protein according to the protocol

10   discussed hereinbefore. The protein fraction was
isolated and analyzed by SDS-PAGE and revealed that the
size of phytoene synthase protein is 38 kilodaltons.

Example 4.   Phytoene Production in E. coli

15   **a.   Method One - Plasmid containing the engineered
genes for GGPP synthase and phytoene synthase**
A plasmid containing genes for both GGPP
synthase and phytoene synthase, as well as an
associated promoter regulatory region adjacent to a

20   structural gene causes E. coli cells containing this
plasmid to produce phytoene. An example of such a
plasmid construct is the use of the structural gene for
GGPP synthase from the plasmid pARC489D with a promoter
that functions in E. coli adjacent to the 5' end of the

25   structural gene for GGPP synthase. This construct is
introduced into a common cloning vector such as pUC18.
Where the structural genes are linked together, a
single promoter can function in E. coli to express both
gene products.

30   A before-described structural gene for
phytoene synthase excised from the plasmid pARC140R is
cloned adjacent to a promoter that functions in
E. coli, such as Rec 7. This Rec 7 promoter-phytoene
synthase heterologous gene is then introduced into a

35   plasmid containing the gene for GGPP synthase. The

plasmid containing both of these genes directs phytoene
synthesis in <u>E. coli</u>. The two genes can also be placed
end-to-end in <u>E. coli</u> under the control of a single
promoter.

5

        b.   **Method Two - Plasmid pARC376 with a defective**
                gene for phytoene dehydrogenase-4H

        Phytoene production can occur with the native
pARC376 plasmid in which the genes for GGPP synthase

10   and phytoene synthase are functional and produce
functional proteins, but in which the gene for phytoene
dehydrogenase-4H is impaired, thereby impairing the
production of lycopene from phytoene. A plasmid
pARC376 derivative in which the gene for phytoene

15   dehydrogenase-4H is deleted or in some other way
impaired could not further metabolize the phytoene
being produced in the <u>E. coli</u> cells due to the action
of the genes for GGPP synthase and phytoene synthase.
Under this condition, phytoene accumulates. The gene

20   for phytoene dehydrogenase-4H is located approximately
between the positions 7849 to 6380 of plasmid pARC376
as shown in Figure 5.

        By example, two different pARC376 derivative
plasmids that contain deletions at the beginning of the

25   gene for phytoene dehydrogenase-4H have been made as
described before. One plasmid is pARC376-Bam 127, in
which the approximately 2749 bp Bam HI fragment from
about position 7775 to about 10524 (Figure 5) was
deleted. The other was plasmid pARC376-Pst 110 missing

30   a Pst fragment at 7792-10791 (Figure 5). These
plasmids were constructed by partially digesting
plasmid pARC376 with either Bam HI or Pst I, and
ligating the respective DNA fragments together.

        These deletions caused the gene for phytoene

35   dehydrogenase-4H to be non-functional, since the

beginning part of the gene was deleted. E. coli cells
that contained either plasmid pARC376-Bam 127 or
plasmid pARC376-Pst 110 produce phytoene. Phytoene is
colorless and cells that produce phytoene have the same
5   colorless character as normal E. coli cells. The
ligation mixture was transformed into E. coli and any
resulting colorless colonies were analyzed for the
presence of phytoene. The presence of phytoene was
confirmed by growing E. coli cells containing the
10   plasmid, performing an extraction according to the
following protocol, and identifying phytoene by HPLC
analysis in the extract.

        c.   Identification of Phytoene Produced by
15   Transformed E. coli
              i.   Extraction from cells
              One hundred to 500 mg of lyophilized E. coli
cells containing an above-described plasmid were
resuspended in 3 ml of 7:2 acetone:methanol in 15 ml
20   conical glass tube with teflon seal cap. 450-600
Micron glass beads (1:1 ratio with the cells) were
added to the tube, which was covered with foil and
vortexed for 2 minutes. After 5 minutes, the tube was
spun in a table top centrifuge and the supernatant
25   transferred to a foil covered glass vial. This
extraction was repeated multiple times.
              The entire pool of the extract was filtered
through a 0.2 micron Acrodisc CR filter in a glass
syringe, and the filtrate was dried under nitrogen.
30   Utmost care was taken to protect the carotenoids/
xanthophylls from light and heat.

        ii.   Identification
              The presence of phytoene was monitored by thin
35   layer chromatography (TLC) analysis in three different

solvent systems using authentic phytoene as a reference.

The carotenoids/xanthophylls were separated by high pressure liquid chromatography (HPLC) with the aid of a Hewlett Packard C-18 Vydac analytical column (4.6 x 250 mm, 5 micron particle size). A linear gradient from 30 percent isopropanol and 70 percent acetonitrile:water (9:1) to 55 percent isopropanol and 45 percent acetonitrile:water (9:1) in 30 minutes (min) at 1 ml/min resolved most of the compounds of interest with the following retention times - zeaxanthin 8.7 min, lycopene 16.2 min, beta-carotene 18.1 min, phytofluene 19.9 min, phytoene 21.8 min, and the zeaxanthin diglucosides were clustered between 6 and 8 min.

The amount of phytoene produced in these cells averaged about 0.01 percent (dry weight).

Example 5. Phytoene Production in S. cerevisiae

S. cerevisiae does not normally produce carotenoids since it does not have the necessary functional genes for phytoene production. S. cerevisiae does, however, produce farnesyl pyrophosphate (FPP). For phytoene production to occur in S. cerevisiae, the genes for GGPP synthase and phytoene synthase need to be transferred into the S. cerevisiae cells in the proper orientation to permit the expression of functional enzymes.

Promoter sequences that function in S. cerevisiae need to be placed adjacent to the 5' end of the structural genes for GGPP synthase and phytoene synthase and termination sequences can also be placed at the 3' ends of the genes. The genes for GGPP synthase and phytoene synthase that contain the proper

-99-

regulatory sequences for expression in <u>S. cerevisiae</u>
then are transferred to the <u>S. cerevisiae</u> cells.

    a.   **Construction of pARC145B**

5   The vector pSOC713 (Figure 8), was made by
first using Klenow polymerase to make blunt ends on the
Eco RI fragment of the yeast B-form 2-micron circle
that contains the 2-micron origin of replication.
Thus, the blunt-ended fragment was cloned into the Sma
10   I site of pUC8. The 2-micron fragment was removed from
the pUC8 construct by cleavage with Eco RI and Bam HI.
This Eco RI-Bam HI fragment was ligated to the Eco RI-
Bgl II fragment of yeast DNA which contains the <u>TRP 1</u>
gene. The DNA containing the fused <u>TRP 1</u> to 2-micron
15   fragment was ligated as an Eco RI fragment into the Eco
RI site of pUC18. Finally, a region of the yeast
genome, containing the divergently-facing <u>GAL 10</u> and
<u>GAL 1</u> promoters was ligated as an Eco RI to Bam HI
fragment into the above <u>TRP</u> 1/2-micron/pUC18 plasmid,
20   which had been cleaved with Eco RI and Bam HI. The
restriction map of plasmid pSOC713 is shown in
Figure 8.
        Three modifications were made to plasmid
pSOC713 to yield plasmid pARC145B (Figure 9). First,
25   plasmid pSOC713 was partially digested with Eco RI and
the ends were made blunt with Klenow polymerase and
self-ligated. The resultant plasmid contained a unique
Eco RI site adjacent to the <u>GAL 1</u> promoter region.
This plasmid was cleaved with Eco RI and the synthetic
30   oligonucleotide shown below,

    5' AATTCCCGGGCCATGGC 3'   (SEQ ID NO:30)
      3' GGGCCCGGTACCGTTAA 5'   (SEQ ID NO:31)

was ligated into the Eco RI site. This regenerated one
35   Eco RI site followed by Sma I and Nco I sites.
Finally, the single Bam HI site was cut, filled in with

-100-

Klenow polymerase, and the Bgl II synthetic linker oligonucleotide

CAGATCTG
GTCTACTG

5   was ligated, cut with Bgl II, and then self-ligated to make a Bgl II site flanked by two Bam HI sites. The restriction map of plasmid pARC145B is shown in Figure 9.

10

b. Construction of Plasmid pARC145G

The engineered gene for GGPP synthase contained in plasmid pARC489D, which encoded the most active version of the enzyme in E. coli above, was

15   transferred to the S. cerevisiae vector pARC145B to generate plasmid pARC145F. This was accomplished by digestion of plasmid pARC489D with Nco I and Pvu II to obtain the approximately 1000 bp Nco I-Pvu II restriction fragment that contained the GGPP synthase

20   structural gene. An Nco I linker was added to the Pvu II site of the restriction fragment to make that fragment an Nco I-Nco I fragment containing about 1010 bp. The GGPP synthase gene was cloned adjacent to the S. cerevisiae divergent promoter region GAL 10 and GAL

25   1 so that the GGPP synthase gene would be expressed in S. cerevisiae using the GAL 10 promoter.

The gene for phytoene synthase from plasmid pARC140R (Example 2) was excised and placed adjacent to the other side of the GAL 1 promoter of plasmid

30   pARC145F so that the phytoene synthase gene would also be expressed using the GAL 1 promoter. Thus, the transcription termination sequence from the S. cerevisiae gene PGK was cloned at the 3' end of the gene for phytoene synthase.

35   More specifically, plasmid pARC145F was digested with Bgl II and Sph I, whose restriction sites

are illustrated in Figure 9 for precursor plasmid
pARC145B. The phytoene synthase structural gene was
excised from plasmid pARC140R as an approximately 1158
Bgl II-Eco RI fragment; the same structural gene is
present in the approximately 1176 bp Hpa I-Eco RI
fragment of plasmid pARC140N. The approximately 500 bp
PGK termination sequence from another plasmid, pARC117,
was excised as an Eco RI-Sph I fragment such as the
same fragment shown in plasmid pARC135 of Figure 7.
The Bgl II-Sph I digested plasmid pARC145F, the Bgl II-
Eco RI about 1158 bp plasmid pARC140R fragment and the
about 500 bp Eco RI-Sph I PGK termination sequence were
triligated to operatively link the three sequences
together.

        This ligation placed the phytoene synthase
structural gene adjacent to and under the control of
the GAL 1 promoter at the 5' end of the structural
gene. The PGK termination sequence was placed at the
3' end of the phytoene synthase structural gene. The
resulting plasmid, now containing both of the genes
required for phytoene production under control of the
GAL 10 and GAL 1 divergent promoters, was named plasmid
pARC145G, and is shown in Figure 10. Other relevant
features of plasmid pARC145G include the 2 micron
origin of replication of S. cerevisiae and the TRP 1
gene of S. cerevisiae as a selectable marker.

        The plasmid pARC145G was transferred into the
S. cerevisiae strain YPH499 (provided by Dr. Phillip
Heiter, Johns Hopkins University) that lacked a
functional TRP 1 gene. This strain was able to utilize
galactose as a carbon source. Transformants were
isolated, and the cells were grown in the presence of
galactose to induce the GAL 10 and GAL 1 promoters to
express the genes for phytoene production.

-102-

The _S. cerevisiae_ cells were grown on the media described below to produce phytoene. YPH499 is a strain of yeast that contains an impaired TRP 1 gene and an impaired URA 3 gene, and is able to utilize galactose as carbon and energy sources. This strain requires tryptophan and uracil in the growth medium in order to grow. Alternatively, these strains can be grown if they are transformed with a plasmid (or plasmids) containing a normal copy of either the TRP 1 gene, but not a normal copy of the URA 3 gene, in which case the cells require uracil to be added to the growth medium, or the URA 3 gene, but not a normal copy of the TRP 1 gene, in which case the cells need to have tryptophan added to the growth medium.

There are four different media used to grow this strain of Saccharomyces:

Medium 1 is used if the cells contain no further URA 3 or TRP 1 genes.

Medium 2 is used if the cells contain a plasmid(s) with only the TRP 1 gene.

Medium 3 is used if the cells contain a plasmid(s) with only the URA 3 gene.

Medium 4 is used if the cells contain a plasmid(s) with both the TRP 1 and the URA 3 genes.

The media constituents are as follows:

Basic Constituents:

0.67% Yeast Nitrogen Base without Amino Acids (Source Difco, #0919-15);

2% Galactose; and

720 mg/l Dropout Mixture*

* Dropout Mixtures

-103-

For Medium 1 (Complete)

Constituent	Amount (mg)
adenine	400
uracil	400
tryptophan	400
histidine	400
arginine	400
methionine	400
tyrosine	600
leucine	1200
lysine	600
phenylalanine	1000
threonine	4000
aspartic acid	2000

For Medium 2, without the tryptophan.

For Medium 3, without the uracil.

For Medium 4, without both tryptophan and uracil.

     To prepare a dropout mixture all of the desired constituents were added to a mortar and ground thoroughly with a pestle. The constituents were thoroughly mixed and 720 mg of the dropout mixture were added for each liter of medium.

     The plasmid pARC145G contains both the GGPP synthase and phytoene synthase genes and a normal copy of the TRP 1 gene. Saccharomyces cells containing pARC145G were grown in Medium 2 with 2 percent galactose.

     The S. cerevisiae cells were analyzed for the presence of phytoene. A total of 0.12 percent (dry weight) phytoene and related compounds having superimposable UV-Vis spectra as phytoene was found in the cells.

-104-

<u>Example 6</u>. **Phytoene Producti n in Higher Plants**

To transfer the genes for GGPP synthase and
phytoene synthase to plants, the <u>Agrobacterium</u> gene
transfer system is preferably used. The structural
5   gene for GGPP synthase discussed before is introduced
into the plasmid pCaMVCN (Pharmacia), by replacing the
structural gene for chloramphenicol acetyltransferase
(<u>CAT</u>) with the gene for GGPP synthase. The <u>Erwinia</u>
<u>herbicola</u> GGPP synthase gene is preferably derived from
10   the <u>E. coli</u> plasmid, pARC489D, described above,
although another <u>Erwinia</u> <u>herbicola</u> GGPP structural gene
can be used.

The GGPP synthase gene in pCaMVCN is adjacent
to the <u>CaMV 35S</u> promoter, and the NOS polyadenylation
15   site is at the 3' end of the GGPP synthase gene. This
gene construct is transferred to the plasmid pGA482
(Pharmacia). The relevant features of the resulting
plasmid are that (i) it contains an origin of
replication that permits it to be maintained in
20   <u>Agrobacterium tumefaciens</u>, (ii) there is a NOS promoter
adjacent to the kanamycin resistance gene that confers
kanamycin resistance to plant cells, (iii) there is a
polycloning site to introduce desired genes to be
transferred to plants, and (iv) the border sequences
25   from the <u>Agrobacterium tumefaciens</u> T-DNA direct the
integration of the desired genes into the plant genome.

The gene for phytoene synthase from plasmids
pARC140R, pARC140N or another, previously described
phytoene synthase structural gene is transferred to
30   pCaMVCN, replacing the <u>CAT</u> gene. This gene is adjacent
to the CaMV 35S promoter, and the NOS polyadenylation
site is at the 3' end of the phytoene synthase gene.
This gene is transferred to the pGA482 derivative that
already contains the gene for GGPP synthase. The
35   result is a gene construct in which both the genes for

GGPP synthase and phytoene synthase can be expressed in
plants.

This vector is transferred to a suitable
Agrobacterium tumefaciens strain such as A281
5   (Pharmacia) or LBA4404 (Clontech). It is noted that
the previously discussed results with A. tumefaciens,
Example lb, illustrate successful introduction of genes
for not only phytoene, but also for additional enzymes
in the carotenoid pathway, the successful expression of
10   the enzymes, and production of phytoene and other
carotenoids in those bacteria.

Subsequently, the genes for GGPP synthase and
phytoene synthase are transferred to the plant genome
after the Agrobacterium tumefaciens cells infect the
15   plant cells. Suitable plants include tobacco and
alfalfa, although others could be used. The genes are
expressed in the growing plant, using the CaMV 35S
promoter, and the enzymes are deposited in the
cytoplasm. Thus, phytoene is produced in the cytoplasm
20   from the engineered enzymes and the naturally
occurring, ubiquitous precursors.

**Example 7.   Phytoene Production in Higher Plant
Chloroplasts**

25   To target the GGPP synthase and phytoene
synthase enzymes to the chloroplast, a DNA sequence
that encodes a transit peptide sequence, which directs
proteins to the chloroplast, is introduced in frame at
the beginning of the genes for these two enzymes. The
30   order of the gene construction, then, is (i) a promoter
that functions in plants, (ii) the DNA sequence for the
transit peptide, (iii) the carotenoid structural gene,
and (iv) a plant polyadenylation sequence. An
exemplary system is discussed below.

-106-

The DNA sequence for the transit peptide from
the "Small Subunit" of the enzyme ribulose bisphosphate
carboxylase from tobacco are synthesized from
oligonucleotide probes. A Nco I site is placed at the
5   5' end of the transit peptide sequence, at the
initiation methionine codon. A Sph I site is placed at
the 3' end of the transit peptide sequence. Details
relating to this transit peptide, its gene and the use
of its gene are in Example 14.

10   The carotenoid genes for GGPP synthase and
phytoene synthase are fused in-frame at the Sph I site
of the transit peptide sequence. The chimeric gene is
cloned into the plasmid pCaMVCN, replacing the CAT
gene.

15   The result of these constructions is a DNA
segment comprising, (i) the plant CaMV 35S promoter
adjacent to the transit peptide sequence, followed by
(ii) a structural gene for either GGPP synthase or
phytoene synthase, or both, and followed by (iii) the
20   NOS polyadenylation site. This gene construct is
transferred to the plasmid pGA482.

The pGA482 plasmid, containing the genes for
GGPP synthase and phytoene synthase, is transferred to
A. tumefaciens. The genes for GGPP synthase and
25   phytoene synthase are transferred into plants following
infection of the plant tissue with the Agrobacterium
strain.

Suitable host plants include tobacco and
alfalfa and others. The genes are expressed from the
30   CaMV 35S promoter, the protein is directed to the
chloroplast by the presence of the transit peptide
sequence, and the enzyme is delivered inside the
chloroplast. Through the action of the engineered
Erwinia herbicola genes for GGPP synthase and phytoene

synthase, and the ubiquitous precursors, phytoene is
accumulated.

Example 8.   Phytoene Dehydrogenase-4H Gen

5   **a.   Localization**

The gene for phytoene dehydrogenase-4H is
found on the plasmid pARC376. The general region of
its location on this plasmid was shown by deleting
specific regions of the pARC376 plasmid and analyzing
10   the carotenoids produced. When an altered or mutated
phytoene dehydrogenase-4H gene is generated, the
phytoene that is produced by the presence of the two
enzymes GGPP synthase and phytoene synthase would
accumulate.

15   The pARC376 plasmid (Figure 5) was partially
digested with either Bam HI or Pst I restriction
enzymes, and the free ends were ligated together. This
DNA was transformed into E. coli HB101, and colorless
colonies were picked and analyzed for the presence of
20   phytoene. Two different plasmid deletions caused the
E. coli cells to accumulate phytoene, including plasmid
pARC376-Bam 127, which had a 2749 bp Bam HI fragment
(7745-10524) deletion and plasmid pARC376-Pst 110,
which had a 2999 bp Pst I fragment (7792-10791)
25   deletion.

The plasmid pARC376-Pst 110 was constructed as
follows. Plasmid pARC376 was partially digested with
Pst I, the DNA was ligated, the ligation mixture was
transformed into E. coli HB101, and the cells were
30   grown in Luria-Broth supplemented with 100 µg/ml
ampicillin. The transformants were screened by
isolating plasmid DNA and performing restriction enzyme
analysis. A plasmid with only the 2999 bp Pst I
segment deleted, was identified and named pARC376-Pst

110. This deletion involves the beginning sequence of
the gene for phytoene dehydrogenase-4H.

        In E. coli cells containing either of these
above two plasmids, phytoene accumulated to about 0.02
5   percent dry weight. This indicated that the gene for
phytoene dehydrogenase-4H was present somewhere in the
deleted region.

        b.   Construction of the plasmid pARC136
10   An about 12,000 bp Eco RI fragment from
plasmid pARC376 was obtained by removal of the segment
from about position 3370 to about position 379 (Figure
5). The resulting large fragment containing all of the
Erwinia herbicola carotenoid genes, was inserted into
15   the Eco RI site of the pBluescript SK + plasmid
(Stratagene, Inc., San Diego) resulting in plasmid
pARC176B. Adjacent to the Eco RI site on the
pBluescript plasmid is a Hind III site. There is
another Hind III site in the insert from plasmid
20   pARC376 (position 13463).

        The plasmid pARC176B was digested with Hind
III, releasing an about 10,200 bp fragment that
contains all of the carotenoid genes. This fragment
was cloned into the Hind III site of the plasmid
25   pARC306A (described before and shown in Figure 6). The
resulting plasmid was named pARC137B.

        There are two Sac I sites in the plasmid
pARC137B; one in the polylinker from plasmid pARC306A,
the other in the GGPP synthase structural gene at about
30   position 11776 (Figure 5). Diagrammatically, the
orientation is as follows:

```
Nco Sac    Hind         Sac                                  Hind
 I   I     III           I                                   III
|---|------|----------|------------------||-----------|
           (13463)    (11776)
```

        The plasmid pARC137B was digested with Sac I,
deleting a 1700 bp Sac I fragment from the Sac I site
in the polylinker to the Sac I site at position 11776.
The remaining large DNA fragment was ligated together,
forming plasmid pARC136, which was transformed into E.
coli HB101, and grown in Luria-Broth supplemented with
100 μg/ml of ampicillin.

        E. coli cells containing plasmid pARC136 were
treated with nalidixic acid to induce the Rec 7
promoter (as described before). One of the proteins
produced was a 51 kilodalton protein, which upon
examination by polyacrylamide gel electrophoresis
(PAGE) was determined to be the phytoene dehydrogenase-
4H enzyme.

        This protein was electroeluted and subjected to
N-terminal amino acid sequencing to obtain the sequence of
the first 30 amino acid residues. Comparison of the
determined amino acid sequence of this 51 kilodalton
protein with the DNA sequence of plasmid pARC376 indicated
that the initiation site of the phytoene dehydrogenase-4H
structural gene is located at about position 7849 of
plasmid pARC376 (Figure 5).

        The 3' end of the phytoene dehydrogenase-4H
gene extends beyond the Bgl II site at position 6836
(Figure 5). The Bgl II site of the insert to plasmid
pARC136 was digested and the ends were polished with
the Klenow fragment of DNA Polymerase I, religated and
transformed into E. coli cells. These manipulations
caused an inhibition of phytoene dehydrogenase-4H and
caused the E. coli cells to accumulate phytoene,

-110-

indicating that the 3' end of the phytoene
dehydrogenase-4H structural gene is downstream from the
Bgl II site.

5   **c.   Construction of the plasmid pARC496A**

The plasmid pARC376 was digested with Sal I
restriction enzyme to excise two adjacent DNA segments;
an about 1092 bp Sal I segment (positions 9340-10432 of
Figure 5), and an about 3831 bp Sal I segment
10   (positions 10432-14263 of Figure 5). The free ends of
the remaining DNA fragment were religated to form the
plasmid, pARC271D.

To introduce a Nco I site at the initiation
methionine of the structural gene for phytoene
15   dehydrogenase-4H, an about 3035 bp Sal I (9340) to Xmn
I (6305 of Figure 5) fragment was excised from plasmid
pARC271D. This fragment was isolated on agarose gel
electrophoresis and used as the template for polymerase
chain reaction (PCR). The following oligonucleotide
20   probe was used:

                  <u>Nco I</u>
         5' AAA CCA TGG AAA AAA CCG TTG TGA TTG GC 3'
                                    (SEQ ID NO:32)
25
For the PCR to run properly, the 3' end must
also be amplified in order to make the proper strands
of the DNA fragment desired. The 3' end of the second
strand oligonucleotide probe retaining the native DNA
30   sequence was:
                  <u>Nco I</u>   (Nco I site at position 6342 of Figure 5)
           5' GG C CAT GG T CTG CGT GGC GTG 3'
                                    (SEQ ID NO:33)
         d.   PCR Reaction:
35   The GeneAmp DNA Amplification Reagent Kit
(Perkin Elmer Cetus) was used to perform the
reaction. The following components were mixed in the

-111-

quantity and order specified according to the
manufacturers instructions.

Component	Order of Addition	Volume (µl)	Final Concentration
Sterile Water	1	43.5	
10 X Rxn. Buffer	2	10	1 X
1.25 mM dNTP Mix	3	16	200 µM @NTP
Primer 1 (10 pMole/µl)	4	10	1 µM
Primer 2 (10 pMole/µl)	5	10	1 µM
Template DNA	6	10	100 ng
Taq Polymerase	7	0.5	2.5 Units

100 µl of mineral oil was layered on top of the
reaction mixture, and the reaction was performed
using the Perkin Elmer Cetus DNA Thermal Cycler
(Perkin Elmer, Prairie Cloud, MN). The method
consisted of 25 cycles of amplification. One cycle
included the following:

1) 1 minute denaturation at 92°C;
2) 2 minute template priming at 37°C;
3) 3 minute polymerization at 72°C; and
4) 7 minute polymerization at 72°C.

After the reaction was completed the mineral
oil was removed, the reaction mixture was extracted
twice with ether, and the DNA was precipitated with
ethanol.

e.   **Cloning of the PCR produced DNA fragment**

1) The DNA produced by the PCR reaction was digested
   with Nco I. This produced a DNA fragment of about
   1509 bp, which was isolated and recovered from an
   agarose gel.

-112-

2) About 5 μg of the plasmid pARC306A was digested
   with Nco I.

3) About 100 ng of the Nco I-digested plasmid
   pARC306A was admixed with about 200 ng of the Nco
5 I fragment produced by the PCR reaction. The
   fragments were inserted using ligation buffer (2
   μl) (IBI Corp.) and 1 Unit of T4 ligase in a total
   volume of 20 μl. The ligation reaction was
   incubated at 4°C for about 15 hours.

10 4) The ligation mixture was transformed into <u>E. coli</u>
   HB101. Transformants were selected on Luria-Broth
   with 100 μg/ml ampicillin. DNA was isolated from
   prospective clones and the clone carrying the
   phytoene dehydrogenase-4H gene insert was
15 identified by restriction enzyme analysis. This
   plasmid was named pARC496A.

        The DNA sequence for the phytoene
dehydrogenase-4H gene was determined as described
20 before and is shown in Figure 11, along with some of
the restriction sites. The approximately 1505 bp Nco
I-Nco I fragment (Nco I fragment) present in plasmid
pARC496A is a particularly preferred DNA segment
herein.

25

    f.   Proof of a Functional Genetically Engineered
          Phytoene Dehydrogenase-4H Gene
        The proper functioning of the gene for
phytoene dehydrogenase-4H in plasmid pARC496A was
30 established by complementation of the plasmid pARC275
(described in Example 9). This plasmid has three
relevant features: i) it is a derivative of plasmid
pARC376 in which part of the gene for phytoene
dehydrogenase-4H gene has been deleted, therefore,
35 the plasmid causes the accumulation of phytoene in <u>E.</u>

-113-

coli; ii) it contains the R1162 origin of
replication; and iii) it contains a kanamycin
resistance gene from Tn5, and therefore, E. coli
cells that contain plasmid pARC275 are able to grow
5   in the presence of 25 µg/ml kanamycin.

E. coli cells containing pARC275 were
transformed with the plasmid pARC496A to form doubly
transformed host cells. These host cells were grown
in medium supplemented with 25 µg/ml kanamycin and
10   100 µg/ml of ampicillin. The cells produced lycopene
at a level of about 0.01 percent dry weight.

This result demonstrated that the gene for
phytoene dehydrogenase-4H had been successfully
engineered. In addition, this result showed that the
15   approximately 1505 bp Nco I-Nco I DNA segment present
in plasmid pARC496A contained the entire DNA sequence
required to produce a functional phytoene
dehydrogenase-4H enzyme.

Because of the introduction of a Nco I site
20   at the initiation methionine of the gene, the
nucleotide sequence was slightly changed:

Original sequence:
                                              (SEQ ID NO:34)
25   5' TAA AGG ATG AAA AAA ACC GTT GTG ATT GGC 3'
                MET Lys Lys Thr Val Val Ile Gly
                                              (SEQ ID NO:35

30   New Genetically Engineered Sequence:

             Nco I   (SEQ ID NO:36)
       5' TAA ACC ATG GAA AAA ACC GTT GTG ATT GGC 3'
                MET Glu Lys Thr Val Val Ile Gly
35   (SEQ ID NO:37)

The sequence at the 3' end of the gene was
not changed as a result of the PCR reaction.

40

-114-

g. Phyt ne D hydrogenase-4H Assay

The assay for phytoene dehydrogenase-4H was developed using two R. sphaeroides mutants, I-3 and E-7. I-3, a mutant strain that has a mutation in the gene for phytoene dehydrogenase-3H, was provided by Dr. Samuel Kaplan, University of Texas Medical Center, Houston, Texas. This mutant, which accumulates phytoene, was used as a source of the substrate for phytoene dehydrogenase-3H and phytoene dehydrogenase-4H.

R. sphaeroides E-7 is a strain that cannot make any carotenoids, and was developed at the Amoco Research Center, Naperville, Illinois. This mutant, which has an intact gene for a different, but similar phytoene dehydrogenase-3H, provided a source of the similar enzyme to determine the proper assay conditions.

The membrane fraction from the Rhodobacter I-3 mutant was isolated by growing I-3 cells until mid to late log phase, pelleting and lysing the harvested cells in 100 mM Tris Buffer, pH 8.0, by vortexing with 150 micron acid-washed glass beads. The cell homogenate was then used as the source of phytoene.

Although the R. sphaeroides E-7 phytoene dehydrogenase-3H transforms phytoene to either phytofluene or neurosporene but not to lycopene, as in Erwinia herbicola, the assay conditions delineated for the Rhodobacter enzyme were also efficacious for the Erwinia herbicola phytoene dehydrogenase-4H. These conditions were used to detect phytoene dehydrogenase-4H activity in both E. coli and S. cerevisiae harboring the Erwinia herbicola structural gene for phytoene dehydrogenase-4H, as is discussed below.

-115-

To isolate the phytoene dehydrogenase-4H from
either bacteria or yeast harboring the Erwinia
herbicola gene, cells were grown until mid-late log
phase and harvested by pelleting. The cell pellet
5   was either frozen for later use or used immediately.
A frozen or fresh cell pellet was resuspended in one
volume of 100 mM Tris Buffer, pH 8.0, and lysed by
vortexing as described above for Rhodobacter (150
micron beads were used to lyse bacteria and 450
10   micron beads were used to lyse yeast). This cell
lysate provided a source of phytoene dehydrogenase-4H
for testing.
An aliquot of the Erwinia herbicola phytoene
dehydrogenase-4H-containing lysate was admixed with
15   an aliquot of the Rhodobacter I-3 cell lysate
described above in a buffer containing 100 mM Tris,
pH 8.0, 10 mM ATP, 2.5 mM NADP, 4 mM DTT, 4 mM MgCl,
6 mM MnCl in a total volume of 1-2 ml. The reaction
mixture was incubated at 30°C in the dark for 2-8
20   hours, and the contents were extracted first with
hexane and then with chloroform. The organic layers
were pooled, dried, and analyzed by HPLC on a C-18
analytical column (4.6 X 250 mm) developed with a
linear gradient, starting with 30 percent isopropyl
25   alcohol and 70 percent acetonitrile:water (9:1) and
ending with 55 percent isopropyl alcohol and 45
percent acetonitrile:water (9:1), in 30 minutes at a
flow rate of 1 ml/minute. Lycopene, which eluted at
about 16.2 minutes, was quantitated from a
30   predetermined standard curve.

-116-

Exampl 9. Lycop ne Production in E. c li
     a.   Method On - Plasmid(s) containing the
           ngin red g n s for GGPP synthas , phytoen
           synthase and phytoene dehydrogenase

Active GGPP synthase, phytoene synthase, and
phytoene dehydrogenase-4H enzymes can convert
ubiquitous cellular precursors into lycopene.
Lycopene was produced in E. coli when plasmids
containing the three genes for the above enzymes were
introduced into the bacterial host cells. One
combination producing lycopene utilized host cells
transformed with the plasmids pARC275 and pARC496A.

The plasmid pARC275 was constructed in the
following manner. First, the plasmid pARC376-Pst 110
was made by deleting the about 2999 bp Pst I segment
(between positions 7792 and 10791, Figure 5) from
plasmid pARC376 as described before. Second, the Eco
RI (3370) to Hind III (13463 Figure 5) segment from
plasmid pARC376-Pst 110 was excised and cloned into
the Eco RI to Hind III sites of plasmid pSOC925 to
produce plasmid pARC275.

The plasmid pSOC925 is about a 9 kilobase
plasmid whose restriction map is illustrated in
Figure 12. This plasmid contains the kanamycin and
chloramphenicol (CAT) resistance genes and the R1162
origin of replication. The chloramphenicol
resistance gene can be excised from the plasmid by
digestion with Eco RI and Hind III (Figure 12).

The fragment (Eco RI to Hind III of plasmid
pARC376-Pst 110) containing the relevant portion of
the Erwinia herbicola carotenoid genes was isolated.
Plasmid pSOC925 was digested with Eco RI and Hind
III, excising the CAT gene. About 100 ng of the
larger portion of digested plasmid pSOC925 was
admixed with about 200 ng of the Eco RI to Hind III

-117-

fragment from pARC376-Pst 110 in a total volume of 20
μl to which 2 μl of Ligation Buffer and 1 Unit of T4
Ligase were added. The ligation mixture was
incubated at 4°C for about 15 hours and then
5   transformed into E. coli HB101 cells. Transformants
were grown in Luria-Broth supplemented with 25 μg/ml
of kanamycin. DNA was isolated from prospective
clones and those clones containing the desired DNA
insert were identified by restriction analysis. The
10   resultant pARC275 plasmid confers the ability to
produce phytoene on E. coli.

Transformation of E. coli host cells with
plasmids pARC275 and pARC496A produced red colonies
of the transformed host cells, as is discussed in
15   Example 8.

b.   Method Two - Plasmid with a defective gene
      for lycopene cyclase
Following production of lycopene, the next
20   step in the Erwinia herbicola biosynthetic pathway is
the transformation of lycopene to beta-carotene by
lycopene cyclase. When the gene encoding lycopene
cyclase is inhibited, mutated, or in some other
manner made non-functional, the enzyme lycopene
25   cyclase, which transforms lycopene to beta-carotene,
does not function. Lycopene accumulates when this
occurs.

The plasmid pARC376-Ava 102, a derivative of
plasmid pARC376 in which the gene for lycopene
30   cyclase has been deleted, was constructed by
partially digesting plasmid pARC376 with Ava I to
remove two adjacent, relatively short Ava I-Ava I
fragments and relegating the cut ends of the
remaining, relatively large fragment. The two
35   relatively small Ava I-Ava I fragments included the

-118-

about 1633 bp Ava I fragment (10453-8842 Figure 5)
and the about 611 bp Ava I-Ava I fragment from (8842-
8231 Figure 5). In total, about 2222 bp of DNA were
deleted from the plasmid pARC376.

5   The resulting plasmid pARC376-Ava 102 was
transformed into E. coli HB101, and the transformants
were grown on Luria-Broth with 100 μg/ml of
ampicillin. Normally, E. coli cells that contain the
entire plasmid pARC376 are yellow due to the

10   production of zeaxanthin and zeaxanthin derivatives.
Following transformation, some of the clones were now
red in color.

          Plasmid DNA was isolated from one of these
red E. coli clones and subjected to restriction

15   analysis, which revealed that the two Ava I-Ava I
fragments had been deleted from the original pARC376
plasmid. This deletion of the Ava I fragments from
plasmid pARC376 impaired the gene for lycopene
cyclase.

20   Under this circumstance, the three genes for
GGPP synthase, phytoene synthase, and phytoene
dehydrogenase-4H on plasmid pARC376-Ava 102
functioned properly and produced lycopene. Because
the gene for lycopene cyclase did not function

25   properly, the transformed E. coli host cells
accumulated lycopene.

Example 10.   Lycopene Production in S. cerevisiae

          Normal yeast cells do not produce lycopene.

30   Genes sufficient to make lycopene in S. cerevisiae
include those for GGPP synthase, phytoene synthase,
and phytoene dehydrogenase-4H. The plasmid pARC145G
(Example 5) has the genes for GGPP synthase and
phytoene synthase on both sides and adjacent to the

35   GAL 10 and GAL 1 divergent promoter region. Both of

these genes are expr ssed in <u>S. cerevisiae</u> using
these two promoters.

    The gene for phytoene dehydrogenase-4H is
located on the plasmid pARC146D described
5   hereinafter. These two plasmids were transformed
into <u>S. cerevisiae</u>, strain YPH499.

    The yeast strain YPH499 contains a non-
functional <u>TRP 1</u> gene and a non-functional <u>URA 3</u> gene
(as discussed in Example 5). Plasmid pARC145G
10 contains a functioning <u>TRP 1</u> gene as well as the
genes for GGPP synthase and phytoene synthase.
Plasmid pARC146D contains a functioning <u>URA 3</u> gene as
well as the gene for phytoene dehydrogenase-4H.
After both plasmids were introduced, the yeast cells
15 were grown on Medium 4 (Example 5) with galactose to
induce the expression of the three carotenoid genes.

    The cells were grown to stationary phase,
collected, extracted, and analyzed by HPLC according
to the protocols described before. Yeast cells with
20 the three carotenoid structural genes produced
lycopene at about 0.01 percent dry weight.

    a.   **Construction of Plasmid pARC146**

    The plasmid pARC146 is a <u>S. cerevisiae</u> vector
25 constructed to direct the expression of the phytoene
dehydrogenase-4H gene in yeast.

    The construction of plasmid pARC145B (Figure
9) was outlined before in Example 5 for production of
phytoene. Two modifications were made to plasmid
30 pARC145B in order to construct plasmid pARC146.

    The first modification was the introduction
of the <u>PGK</u> terminator at the Sph I site of plasmid
pARC145B, downstream from the <u>GAL 1</u> promoter. A
polycloning site, into which a structural gene could

be cloned, is present between the GAL 1 promoter and
the PGK terminator.

Thus, an about 500 bp Eco RI-Hind III
fragment containing the S. cerevisiae PGK terminator
5   was excised from plasmid pARC117 (Example 5). This
is substantially the same PGK terminator fragment
discussed in Example 5 and shown in Figure 7 for
plasmid pARC135. The Eco RI and Hind III ends of
this fragment were blunted by treatment with the
10   Klenow fragment of DNA Polymerase. Synthetic double-
stranded sequences each containing a potential Sph I
cleavage site (BRL) were then ligated to both ends of
the PGK terminator fragment, and that fragment was
digested with Sph I, producing sticky ends. Plasmid
15   pARC145B was digested with Sph I, and the Sph
I-linkered PGK terminator was ligated to form the
resulting plasmid pARC145C.

The second modification was to replace the
yeast TRP 1 gene with the yeast URA 3 gene. This
20   enabled transfer of the plasmid into yeast cells that
had a mutation in the URA 3 gene on the yeast
chromosome. Here, the plasmid pARC145C was digested
with restriction enzymes Msc I and Eco RV, and a 737
bp fragment containing the TRP 1 gene was deleted.
25   Synthetic double-stranded sequences
containing a potential Xho I cleavage site (BRL) were
ligated to the Msc I and Eco RV blunt ends (there are
no other Xho I sites in plasmid pARC145). The
resulting DNA fragment was digested with Xho I to
30   produce a DNA having Xho I sticky ends.

Meanwhile, an about 1000 bp Hind III
fragment, including the entire URA 3 gene, was
excised from the plasmid YEp24 (ATCC 37051). The
ends of this fragment were blunted with the Klenow
35   fragment of DNA Polymerase. Synthetic double-

stranded sequences, each containing a potential Xho I
cleavage site were ligated to the blunt ends. This
fragment was then digested with Xho I, producing
sticky ends.

5   This URA 3 gene fragment was then ligated
into the Xho I-digested pARC145C plasmid (from which
the TRP 1 gene had been deleted). The final plasmid
was named pARC146 and is similar to plasmid pARC145C
except that plasmid pARC146 contains a URA 3

10   selectable maker instead of a TRP 1 gene.
           Unexpectedly, plasmid pARC146 did not contain
two Xho I sites. The Xho I site expected at the
location of the Eco RV site of the original vector,
denoted as (Xho I) in Figure 13, could not be

15   digested. However, the apparent loss of the site did
not effect the utility of plasmid pARC146 as a URA 3
selectable vector and also did not effect the utility
of plasmid pARC146 as an expression vector.
           The relevant features of this new plasmid

20   construct are i) the presence of the divergent GAL 1
and GAL 10 promoters, ii) the PGK terminator at the
3' end of the GAL 1 promoter, iii) the 2 micron STB
terminator (2 MIC STB TERM) at the 3' end of the GAL
10 promoter, iv) the URA 3 gene that is the

25   selectable marker for transferring the plasmid into
S. cerevisiae, and v) the 2 micron origin of
replication that permits the maintenance of the
plasmid in yeast. This plasmid also contains the
pMB1 origin of replication for maintenance in E. coli

30   and the ampicillin resistance gene for selection in
E. coli. A restriction map of the plasmid pARC146 is
shown in Figure 13.

-122-

b.   C nstructi n of Plasmid pARC496B

Plasmid pARC496B was constructed to introduce
a Sal I site immediately upstream from the initiation
methionine of the phytoene dehydrogenase-4H

5   structural gene and a Sal I site at the 3' end of the
gene to enable the gene for phytoene dehydrogenase-4H
to be moved as a Sal I-Sal I fragment. This version
of the gene was used as the structural gene for
phytoene dehydrogenase-4H in constructing the plasmid

10   pARC146D (described below) that was transformed into
S. cerevisiae in combination with transformation with
plasmid pARC145G to cause the production of lycopene
in the transformed yeast. The plasmid pARC496B was
constructed using the PCR protocol described before

15   (plasmid pARC496A) to introduce Sal I sites at the 5'
and 3' end of the gene.

i.   Template DNA for the PCR

The plasmid pARC271D (Example 8) was digested

20   with Sal I and Xmn I and an about 3035 bp fragment
(9340-6305, Figure 5) was isolated after separation
on agarose gel electrophoresis. This fragment was
used as the template for PCR.

25   **ii.   Probes for the PCR**

Two oligonucleotide probes were used to
introduce Sal I sites at the 5' and the 3' ends of
the gene for phytoene dehydrogenase-4H. At the 5'
end of the gene, the newly introduced Sal I site was

30   immediately upstream from the initiation methionine.
At the 3' end of the gene, the newly introduced Sal I
site was immediately upstream from the Nco I site at
6342.

35

The original sequence of the 5' end was:
5'   G AGA TAA AGG ATG AAA AAA ACC GTT GTG AT   3'
                         MET ...   (SEQ ID NO:38)

The oligonucleotide probe for the 5' end was:

                   <u>Sal I</u>
5'   G AGG TCG ACG ATG AAA AAA ACC GTT GTG AT   3'
                         MET ...   (SEQ ID NO:39)
The second strand oligonucleotide probe for the 3'
end of the gene was:
           <u>Sal I</u>   (SEQ ID NO:40)
5'   AT GGT CGA CGT GGC GTG GTC AAG CAG CGG   3'

        The polymerase chain reaction was carried out
as described before. After completion, the reaction
mixture was extracted twice with ether and the DNA
was precipitated with ethanol.

        iii. Cloning of the PCR produced DNA Fragment
        The DNA accumulated from the PCR was digested
with Sal I, producing an about 1508 bp fragment (from
the "T" of the TCGAC overhang at the 5' end of the
gene to the "G" of the Sal I site at the 3' end of
the gene). Five μg of the plasmid pARC306A (Figure
6) was digested with Sal I. About 100 ng of the Sal
I-digested plasmid pARC306A and about 200 ng of the
Sal I-Sal I fragment of the phytoene dehydrogenase-4H
structural gene prepared by PCR were admixed with 2
μl of Ligation Buffer (IBI) and 1 Unit of T4 Ligase
in a total volume of 20 μl. The ligation reaction
mixture was incubated at 4°C for about 15 hours.
        The resulting plasmid was transformed into <u>E.</u>
<u>coli</u> HB101, and the transformants were selected by
growth in Luria-Broth supplemented with 100 μg/ml of
ampicillin. DNA from prospective clones was isolated
and the identity of clones containing the phytoene

-124-

dehydrogenase-4H gene was confirmed by restriction enzyme analysis.

The resultant plasmid was named pARC496B. The about 1508 bp Sal I-Sal I fragment (also referred to as a Sal I fragment), another particularly preferred DNA segment herein, was cloned from plasmid pARC496B into the yeast vector pARC146, to generate the plasmid pARC146D as described hereinafter.

iv. Sequence of the Phytoene Dehydrogenase-4H Gene Fragment of Plasmid pARC496B

The introduction of the Sal I sites at the 5' and 3' ends of the gene for phytoene dehydrogenase-4H changed the nucleotide sequence of the native DNA fragment slightly.

Original sequence at the 5' end of the gene:

```
                                       (SEQ ID NO:41)
5' GAG ATA AAG G ATG AAA AAA ACC GTT GTG AT   3'
               MET Lys Lys Thr Val Val ...
                                       (SEQ ID NO:42)
```

Sequence of the genetically engineered versions of the gene at the 5' end:

```
         Nco I           (SEQ ID NO:43)
      5' CC ATG GAA AAA ACC GTT GTG AT   3'
         MET Glu Lys Thr Val Val
                         (SEQ ID NO:44)

         Sal I           (SEQ ID NO:45)
5' GAG GTC GAC G ATG AAA AAA ACC GTT GTG AT   3'
               MET Lys Lys Thr Val Val ...
                         (SEQ ID NO:46)
```

Original sequence at the 3' end of the gene:

```
                               (SEQ ID NO:47)
                          Nco I  (6395)
5' CC GCT GCT TGA CCA CGC CAC GCA GAC CAT GG  3'
```

-125-

After the introduction of the Sal I site from the PCR
reaction the new sequence became:

<pre>
                                       Sal I Nco I (6395)
5' CC GCT GCT TGA CCA CGC CAC GTC GAC CAT GG   3'
                                       (SEQ ID NO:48)
</pre>

c.   Construction of the plasmid pARC146D

An about 1508 bp Sal I fragment described
above containing the structural gene for phytoene
dehydrogenase-4H was excised from plasmid pARC496B
and was ligated into the Sal I site of the pARC146
plasmid described before. The result was the plasmid
pARC146D construct, placing the gene for phytoene
dehydrogenase-4H between and adjacent to the GAL 1
promoter and the PGK terminator. A restriction map
of the pARC146D plasmid is illustrated in Figure 14,
in which the location of the phytoene dehydrogenase-
4H gene is shown as "PDH".

**Example 11.   Expression of Erwinia Herbicola Phytoene
Dehydrogenase-4H Gene in Rhodobacter
sphaeroides**

This Example describes the construction
of a plasmid, pATC228, that was transformed into a
mutant strain of R. sphaeroides, causing the
expression of Erwinia herbicola phytoene
dehydrogenase-4H in that organism. Plasmid vector
pATC228 was made by combining the plasmid pATC1619,
which contains a genetically engineered phytoene
dehydrogenase-4H structural gene, with plasmid
pSOC244, which is capable of transforming and being
maintained in both E. coli and R. sphaeroides. The
following is a description of the multistep
construction of plasmid pATC228.

-126-

a.   **Construction of Plasmid pATC1619**

The plasmid pATC1619 contains a genetically
engineered version of the phytoene dehydrogenase-4H
gene cloned adjacent to the TAC promoter of pDR540
5   (Pharmacia). The gene for phytoene dehydrogenase-4H
is expressed in E. coli and photosynthetic bacteria
using the TAC promoter. Plasmid pATC1619 was
constructed in a multistep procedure requiring
several intermediate plasmids as outlined below.
10

i.   **Plasmid pARCBglII401**

The plasmid pARCBglII401 was constructed by
cloning the about 5513 bp Bgl II fragment from
plasmid pARC376 (from position 6836 to position 12349
15   in Figure 5) into the Bam HI site of plasmid pARC306A
(Figure 6).

ii.   **Plasmid pATC1403**

An about 1548 bp Pst I to Sal I fragment from
plasmid pARCBglII401 (original coordinates in Figure
20   5 were 7792 and 9340, respectively) was cloned into
the Pst I and Sal I sites of plasmid M13mp19 (BRL) to
generate plasmid pARC1403. Plasmid pATC1403 contains
a beginning portion of the phytoene dehydrogenase-4H
gene.
25

iii.   **Plasmid pATC1404**

A Sph I site was introduced at the initiation
MET codon of the phytoene dehydrogenase-4H gene in
plasmid pATC1403, using the in vitro mutagenesis
30   protocol described in Current Protocols in Molecular
Biology, Ausabel et al. eds., John Wiley & Sons, New
York (1987), pp. 8.1.1-8.1.6 (see Example 2). The
oligonucleotide probe used as the primer was:

35

-127-

```
                Sph I              (SEQ ID NO:49)
    5' G ACG AGA TAA AGC ATG CAA AAA ACC GTT GT 3'
                   MET Gln Lys Thr Val ...
                                   (SEQ ID NO:50)
```

The sequence in the native phytoene dehydrogenase-4H
gene was:

```
                          .        (SEQ ID NO:51)
    5' G ACG AGA TAA AGG ATG AAA AAA ACC GTT GT 3'
                   MET Lys Lys Thr Val ...
                          .        (SEQ ID NO:52)
```

As a result of the introduction of the Sph I site,
the second amino acid of the phytoene
dehydrogenase-4H enzyme was changed from Lys to Gln.
Thus, the new sequence became:

```
                Sph I              (SEQ ID NO:53)
    5' G ACG AGA TAA AGC ATG CAA AAA ACC GTT GT 3'
                   MET Gln Lys Thr Val
                                   (SEQ ID NO:54)
```

This plasmid, with the Sph I site at the initiation
methionine codon of the phytoene dehydrogenase-4H
structural gene, was named pATC1404.

iv. Plasmid pATC816

The plasmid, pARC306A (Figure 6) was digested
with Pst I and Sma I. The plasmid pARC376 (Figure 5)
was digested with Pst I and Bal I. An about 1451 bp
Pst I (7792) to Bal I (6341) fragment was isolated
from an agarose gel. Both, Bal I and Sma I
digestions leave a blunt end. The approximately 1451
bp Pst I-Bal I fragment from plasmid pARC376 was
cloned into the Pst I and Sma I digested plasmid
pARC306A to form plasmid pATC816.

Plasmid pARC306A contains an Eco RI site
about 30 bp downstream from the Sma I site. The Eco
RI site originally present in plasmid pARC306A is
maintained in plasmid pATC816.

-128-

v. Plasmid pATC1605

As previously stated, the plasmid pATC1404 contains only the beginning portion of the gene encoding phytoene dehydrogenase-4H. To fuse this portion with the remainder of the phytoene dehydrogenase-4H gene, an about 1052 bp Sma I to Pst I fragment from plasmid pATC1404 (original position 8844 to 7792 of pARC376 in Figure 5) was excised and cloned into plasmid pATC816 (which contains the 3' portion of the phytoene dehydrogenase-4H gene) as follows.

Plasmid pATC816 was digested with Ssp I and Pst I (both sites are unique in the pATC816 plasmid). Digestion with Ssp I left a blunt end. The about 1052 bp Sma I (blunt end) to Pst I fragment from plasmid pATC1404 was cloned into the digested plasmid pATC816, resulting in plasmid pATC1605.

This cloning procedure (Sma I to Pst I fragment of plasmid pATC1404 into plasmid pATC816) resulted in the complete sequence of the phytoene dehydrogenase-4H gene becoming contiguous in plasmid pATC1605. There is a superfluous DNA segment immediately upstream from the initiation codon of the phytoene dehydrogenase-4H gene.

In addition, the newly created Sph I site of plasmid pATC1404 containing the codon for the initial Met residue of the enzyme became a part of the phytoene dehydrogenase-4H structural gene. The originally present Nco I site shown near the 3' end of the sequence of Figure 11-4 is also present in this construct as is the Eco RI site downstream therefrom that was introduced from plasmid pARC306A. The Sph I-Eco RI fragment of plasmid pATC1605 that contains the structural gene for phytoene dehydrogenase-4H contains about 1550 bp.

vi. Plasmid pATC1607

Plasmid pATC1605 was digested with Sph I and Eco RI enzymes. The resultant fragment of about 1550 bp was cloned into the plasmid pUC19 (Pharmacia),

5   which had been digested with Sph I and Eco RI enzymes, resulting in the plasmid, pATC1607.

vii. Plasmid pATC1619

Upstream and adjacent to the Sph I site on

10   plasmid pATC1607 is a Hind III site that originates from the polylinker region of plasmid pUC19. The structural gene for phytoene dehydrogenase-4H was excised from plasmid pATC1607 by digesting with Hind III and Eco RI. The ends of the resultant fragment,

15   also of about 1550 bp, were blunted by treating with the Klenow fragment of E. coli DNA Polymerase.

The plasmid, pDR540 (Pharmacia), which contains the TAC promoter for gene expression in some bacteria, including E. coli and R. sphaeroides, and

20   a unique Bam HI site downstream of the TAC promoter, was digested with Bam HI, and the ends were blunted as above. The blunt ended DNA fragment from plasmid pATC1607 (above) was cloned into plasmid pDR540, resulting in the plasmid pATC1619, which contained

25   the bacterial TAC promoter adjacent to the structural gene for phytoene dehydrogenase-4H. Plasmid pATC1619 also contains a unique Hind III site.

b.   Construction of Plasmid pATC228

30   Plasmid pSOC244 is a plasmid that contains i) the R1162 origin of replication, ii) the chloramphenicol acetyltransferase gene that confers resistance to chloramphenicol adjacent to the TAC promoter, and iii) a unique Hind III site. This

35   plasmid can transform and be maintained in both

E. coli and R. sphaeroides. The construction of
plasmid pSOC244 is discussed below.

i.   Plasmid pSOC200

Plasmid pQR176a was obtained from Dr. J.A.
Shapiro of the University of Chicago, Chicago, IL,
and is described in Meyer et al., J. Bacteriol.,
152:140 (1982). This plasmid contains the R1162
origin of replication and the transposon Tn5, which
confers resistance to kanamycin. This plasmid
contains about 14.5 kilobases and contains several
Hind II restriction sites.
    Digestion of plasmid pQR176a with Hind II,
followed by religation of appropriate fragments
provided plasmid pSOC200, which contained about 8.5
kilobases. This plasmid retained the R1162 origin of
replication and the kanamycin resistance gene from
Tn5.

ii. Plasmid pSOC244

Plasmid pSOC200 was digested with Hind III
and Sma I endonucleases to remove the kanamycin
resistance gene. Plasmid pSOC925 was similarly
digested to provide an approximately 1000 bp fragment
containing the chloramphenicol acetyltransferase
(CAT) structural gene with the adjacent TAC promoter.
That approximately 1000 bp fragment was then cloned
into the Hind III- and Sma I-digested plasmid pSOC200
fragment to provide plasmid pSOC244.

iii. Plasmid pATC228

Both plasmids, pATC1619 and pSOC244, were
digested with Hind III. The two plasmids were
ligated together and selected in E. coli grown in
medium containing ampicillin (using the ampicillin

-131-

resistance gene from the pATC1619 plasmid) and
chloramphenicol (using the chloramphenicol resistance
gene from the pSOC244 plasmid). The resultant
plasmid was pATC228, which contains the structural

5   gene for phytoene dehydrogenase-4H and can transform
and be maintained in R. sphaeroides. This structural
gene can be excised from plasmid pATC228 as an
approximately 1506 bp Sph I-Nco I restriction
fragment. Plasmid pATC228 is shown schematically in

10   Figure 16.

        c.   Expression of the Erwinia Herbicola Phytoene
              Dehydrogenase-4H Gene in a R. sphaeroides I-3
              Mutant

15   The R. sphaeroides I-3 mutant (utilized in
Example 8g), possesses an impaired native crtI gene
for phytoene dehydrogenase-3H, and thus accumulates
phytoene. Cells from R. sphaeroides I-3 were
transformed as hosts with plasmid pATC228. The

20   transformants were selected in the presence of
chloramphenicol. The mutant cells that were
previously colorless, were colored red after
transformation. The red pigment produced by these
cells had physicochemical characteristics that were

25   consistent with the properties of the carotenoid
spirilloxanthin.
              The pigment produced by the plasmid pATC228-
transformed R. sphaeroides I-3 mutant host cells was
compared to authentic spirilloxanthin extracted from

30   R. rubrum (ATCC 25903) cells grown in culture. The
two pigments had the same UV-Vis spectra and the same
HPLC profiles. Spirilloxanthin from R. rubrum is
derived from lycopene through a series of catalytic
steps that include two dehydrogenations, hydration,

35   and then methylation. The Photosynthetic Bacteria,

-132-

Roderick et al. eds., Plenum Press, New York, pages
729-750 (1978).

R. sphaeroides normally transforms phytoene
to neurosporene, but not to lycopene, as is the case
in Erwinia herbicola. It is believed, therefore,
that in the production of the spirilloxanthin-like
pigment in the transformed R. sphaeroides, the
Erwinia herbicola phytoene dehydrogenase-4H catalyzed
desaturation of accumulated phytoene to produce
lycopene. The produced lycopene was thereafter
further metabolized by native enzymes present in the
R. sphaeroides mutant to form spirilloxanthin-like
carotenoid.

Example 12. Lycopene Production in Pichia pastoris
The above-described method is also extendable
to other yeasts. One yeast system that serves as an
example is the methylotrophic yeast, Pichia pastoris.
To produce lycopene in P. pastoris,
structural genes for GGPP synthase, phytoene
synthase, and phytoene dehydrogenase-4H are placed
under the control of regulatory sequences that direct
expression of structural genes in Pichia. The
resultant expression-competent forms of those genes
are introduced into Pichia cells.

For example, the transformation and
expression system described by Cregg et al.,
Biotechnology 5:479-485 (1987); Molecular and
Cellular Biology 12:3376-3385 (1987) can be used. A
structural gene for GGPP synthase such as that from
plasmid pARC489D is placed downstream from the
alcohol oxidase gene (AOX1) promoter and upstream
from the transcription terminator sequence of the
same AOX1 gene. Similarly, structural genes for
phytoene synthase and phytoene dehydrogenase-4H such

-133-

as that from plasmids pARC140N and pARC146D are
placed between AOX1 promoters and terminators. All
three of these genes and their flanking regulatory
regions are then introduced into a plasmid that
5   carries both the P. pastoris HIS4 gene and a P.
pastoris ARS sequence (Autonomously Replicating
Sequence), which permit plasmid replication within P.
pastoris cells [Cregg et al., Molecular and Cellular
Biology, 12:3376-3385 (1987)].

10   The vector also contains appropriate portions
of a plasmid such as pBR322 to permit growth of the
plasmid in E. coli cells. The final resultant
plasmid carrying GGPP synthase, phytoene synthase,
and phytoene dehydrogenase-4H genes, as well as the
15   various additional elements described above, is
illustratively transformed into a his4 mutant of
P. pastoris, i.e. cells of a strain lacking a
functional histidinol dehydrogenase gene.

            After selecting transformant colonies on
20   media lacking histidine, cells are grown on media
lacking histidine, but containing methanol as
described by Cregg et al., Molecular and Cellular
Biology, 12:3376-3385 (1987), to induce the AOX1
promoters. The induced AOX1 promoters cause
25   expression of the enzymes GGPP synthase, phytoene
synthase, and phytoene dehydrogenase-4H and the
production of lycopene in P. pastoris.

            The three genes for GGPP synthase, phytoene
synthase, and phytoene dehydrogenase-4H can also be
30   introduced by integrative transformation, which does
not require the use of an ARS sequence, as described
by Cregg et al., Molecular and Cellular Biology,
12:3376-3385 (1987).

35

-134-

Exampl 13. Lycop n Production in A. nidulans

        The genes encoding GGPP synthase, phytoene
synthase, and phytoene dehydrogenase-4H as discussed
before can be used to synthesize and accumulate
5   lycopene in fungi such as Aspergillus nidulans.
Genes are transferred to Aspergillus by integration.

        For example, the structural gene for GGPP
synthase is introduced into the E. coli plasmid
pBR322. The promoter from a cloned Aspergillus gene
10   such as argB [Upshall et al., Mol. Gen. Genet.
204:349-354 (1986)] is placed into the plasmid
adjacent to the GGPP synthase structural gene. Thus,
the GGPP synthase gene is now under the control of
the Aspergillus argB promoter.

15   Next, the entire cloned amds gene [Corrick
et al., Gene 53:63-71 (1987)] is introduced into the
plasmid. The presence of the amds gene permits
acetamide to be used as a sole carbon or nitrogen
source, thus providing a means for selecting those
20   Aspergillus cells that have become stably transformed
with the amds-containing plasmid.

        Thus, the plasmid so prepared contains the
Aspergillus argB promoter fused to the GGPP synthase
gene and the amds gene present for selection of
25   Aspergillus transformants. Aspergillus is then
transformed with this plasmid according to the method
of Ballance et al., Biochem. Biophys. Res. Commun.
112:284-289 (1983).

        The GGPP synthase, phytoene synthase and
30   phytoene dehydrogenase-4H structural genes are each
similarly introduced into the E. coli plasmid pBR322.
Promoters for the cloned Aspergillus argB gene
[Upshall et al., Mol. Gen. Genet, 204:349-354 (1986)]
are placed immediately adjacent to the phytoene
35   synthase and phytoene dehydrogenase-4H structural

genes. Thus, these structural genes are controlled
by the Aspergillus argB promoters.

The entire, cloned Aspergillus trpC gene
[Hamer and Timberlake, Mol. Cell. Biol., 7:2352-2359
5   (1987)] is introduced into the plasmid. The trpC
gene permits selection of the integrated plasmid by
virtue of permitting transformed trpC mutant
Aspergillus cells to now grow in the absence of
tryptophan. The Aspergillus strain, already
10   transformed with the plasmid containing the GGPP
synthase gene, is now capable of synthesizing
lycopene.

Example 14. Phytoene Dehydrogenase-4H in Higher
15   **Plants**
Higher plants have the genes encoding the
enzymes required for lycopene production and so
inherently have the ability to produce lycopene.
Lycopene normally is not accumulated, however,
20   because lycopene so produced in most plants is
further converted to other products. Even in the
case of ripe tomato fruits, the level of lycopene
accumulated is only about 0.01 percent dry weight.
The carotenoid-specific genes from Erwinia herbicola
25   can be used to express phytoene dehydrogenase-4H for
use by the plant as well as to improve accumulation
of lycopene in plants. Two useful approaches are
described below.

30   **a.   Transport to the Chloroplast**
In the first approach, the gene for phytoene
dehydrogenase-4H was modified to introduce the
restriction site Sph I at the initiation methionine
codon, as discussed before. An about 177 bp DNA
35   fragment that encodes for the transit (signal)

peptide of the tobacco gene for ribulose bis-
phosphate carboxylase-oxygenase containing a Nco I
site at the 5' end and a Sph I site at the 3' end,
was ligated to the Sph I site of the structural

5   phytoene dehydrogenase-4H gene. This modified gene
was inserted into the plasmid pCaMVCN (Pharmacia,
Piscataway, N.J.) replacing the CAT gene. The
resultant plasmid contained a gene for phytoene
dehydrogenase-4H with the transit peptide sequence

10 placed between and adjacent to both the CaMV 35S
plant promoter and the NOS polyadenylation sequence
at the 3' end.

        This phytoene dehydrogenase-4H gene construct
was inserted into the plasmid pGA482 (Pharmacia) in a

15 convenient restriction site within the multiple
cloning linker region to form plasmid pATC1616. The
relevant features of plasmid pGA482 include (i) an
origin of replication that permits maintenance of the
plasmid in Agrobacterium tumefaciens, (ii) the left

20 and right border sequences from the T-DNA region that
direct the integration of the DNA segment between the
borders into the plant genome, and (iii) the NOS
promoter adjacent to the kanamycin resistance gene
that permits plant cells to survive in the presence

25 of kanamycin.

        This phytoene dehydrogenase-4H gene construct
was transformed into Agrobacterium tumefaciens
LBA4404 (Clontech, Inc.) according to standard
protocols. Agrobacterium cells containing the

30 plasmid with the phytoene dehydrogenase-4H gene
construct were transferred by infection of tobacco
leaf discs using the method of Horsch et al.,
Science, 227:1229-1231 (1985). During the infection
process, the entire DNA segment between the left and

35 right borders of the plasmid pGA482 plasmid is

-137-

transfected into the plant cells. Transfected plant
cells are selected for kanamycin resistance.

Transgenic tobacco plants were grown in the
presence of the herbicide norflurazon (Sandoz).

5   Control plants that had been transformed with the
control plasmid pGA482 and that did not contain
Erwinia herbicola phytoene dehydrogenase-4H
structural gene bleached when grown in the presence
of 0.2 μg/ml in the growth medium. Transgenic plants

10   containing the Erwinia herbicola phytoene
dehydrogenase-4H structural gene grew normally in the
presence of 0.8 μg/ml of Norflurazon. Thus, the
introduction of the Erwinia herbicola phytoene
dehydrogenase-4H structural gene caused the

15   expression of Erwinia herbicola phytoene
dehydrogenase-4H, and plants to become resistant to a
herbicidal amount of norflurazon.

The specific DNA segments, recombinant
molecules and techniques utilized in the preparation

20   of the above norflurazon-resistant tobacco plants are
discussed below.

        i.   Transit Peptide

            The sequence of the transit peptide DNA

25   is basically that of Mazur et al., Nucl. Acids Res.,
13:2343-2386 (1985) for the ribulose bis-phosphate
carboxylase-oxygenase signal peptide of Nicotiana
tabacum. Two changes were made to the disclosed 177
bp sequence.

30   In the first change, two cytidine residues
were added at the 5' end to create a Nco I
restriction site. The second change introduced an
Nar I site that cleaves between bases at positive 73
and 74. This change was a G for T replacement at

35   position 69 and a G for A replacement at position 72,

-138-

both of which changes left the encoded amino acid
residue sequence unchanged. The final two residues
at the 3' end were deleted to provide the natural Sph
I restriction site sticky end.

5   The synthetic transit peptide-encoding DNA
also therefore contained 177 bp. The complete double
stranded sequence, showing the 5' Nco I and 3' Sph I
sticky ends, is illustrated in Figure 17.
           The DNA encoding the transit peptide was
10   synthesized synthetically from eight fragments that
were annealed together in pairs by heating at 90°C
for five minutes and then slowly cooling to room
temperature. Fifty picomoles of each fragment were
utilized.

15   Those eight fragments were:

1. 5' CAT GGC TTC CTC AGT TCT TTC CTC TGC AGC AGT
      TGC C 3'   (SEQ ID NO:55)

2. 5' GGG TGG CAA CTG CTG CAG AGG AAA GAA CTG AGG
20   AAG C 3'   (SEQ ID NO:56)

3. 5' ACC CGC AGC AAT GTT GCT CAA GCT AAC ATG
      GTG G 3'   (SEQ ID NO:57)

25 4. 5' CGC CAC CAT GTT AGC TTG AGC AAC ATT GCT GC 3'
                                  (SEQ ID NO:58)

5. 5' CGC CTT TCA CTG GCC TTA AGT CAG CTG CCT CAT
      TCC CTG TTT CAA GGA AG 3'   (SEQ ID NO:59)

30
6. 5' TTT GCT TCC TTG AAA CAG GGA ATG AGG CAG CGA
      ATG AGG CAG CTG ACT TAA GGC CAG TCA AAG G 3'
                                  (SEQ ID NO:60)

35

-139-

7. 5' CAA AAC CTT GAC ATC ACT TCC ATT GCC AGC AAC
        GGC GGA AGA GTG CAA TGC ATG 3'

                                        (SEQ ID NO:61)

8. 5' CAT TGC ACT CTT CCG CCG TTG CTG GCA ATG GAA
        GTG ATG TCA AGG T 3'   (SEQ ID NO:62)

        The pairs utilized for annealing were 1 and
2, 3 and 4, 5 and 6, and 7 and 8 to form sticky ended
annealed pairs 1-2, 3-4, 5-6 and 7-8 that are shown
below.
1-2

                                        (SEQ ID NO:63)
5' CATGGCTTCCTCAGTTCTTTCCTCTGCAGCAGTTGCC 3'
    3'CGAAGGAGTCAAGAAAGGAGACGTCGTCAACGGTGGG 5'
                                        (SEQ ID NO:64)
3-4

                                        (SEQ ID NO:65)
5' ACCCGCAGCAATGTTGCTCAAGCTAACATGGTGG 3'
        3'CGTCGTTACAACGAGTTCGATTGTACCACCGC 5'
                                        (SEQ ID NO:66)
5-6

5' CGCCTTTCACTGGCCTTAAGTCAGCTGCCTCATTCCCTGTTTCA
    3' GGAAAGTGACCGGAATTCAGTCGACGGAGTAAGGGACAAAGT

        AGGAAG 3'   ·(SEQ ID NO:67)
        TCCTTCGTTT 5'   (SEQ ID NO:68)

7-8

5' CAAAACCTTGACATCACTTCCATTGCCAGCAACGGCGGAAGAGT
    3' TGGAACTGTAGTGAAGGTAACGGTCGTTGCCGCCTTCTCA

        GCAATGCATG 3'   (SEQ ID NO:69)
        CGTTAC 5'   (SEQ ID NO:70)

        Fragment 1-2 was ligated with fragment 3-4 to
form fragment 1-4 whose sequence is shown below.

5' CATGGCTTCCTCAGTTCTTTCCTCTGCAGCAGTTGCCACCCGCAGCAA
    3' CGAAGGAGTCAAGAAAGGAGACGTCGTCAACGGTGGGCGTCGTT

        TGTTGCTCAAGCTAACATGGTGG 3'   (SEQ ID NO:71)
        ACAACGAGTTCGATTGTACCACCGC 5'   (SEQ ID NO:72)

-140-

Fragment 5-6 was ligated with fragment 7-8 to form fragment 5-8 whose sequence is shown below.

5   5' CGCCTTTCACTGGCCTTAAGTCAGCTGCCTCATTCCCTGTTTCAAGGA
       3' GGAAAGTGACCGGAATTCAGTCGACGGAGTAAGGGACAAAGTTCCT

       AGCAAAACCTTGACATCACTTCCATTGCCAGCAACGGCGGAAG
       TCGTTTTGGAACTGTAGTGAAGGTAACGGTCGTTGCCGCCTTC

10   AGTGCAATGCATG 3'   (SEQ ID NO:73)
       5'TCACGTTAC   (SEQ ID NO:74)

The 1-2 and 3-4 pairs (fragments 1-4) were
15   ligated together over a two hour time period, as were
pairs 5-6 and 7-8 to form two double-stranded
sequences. The ligation product of fragments 1-4 was
digested with Nco I and Nar I, whereas the product of
fragments 5-8 was digested with Nar I and Sph I.
20   These digestions separated any concatamers formed
during ligation and provided the necessary sticky
ends for further ligation.

The digested mixes were run on 6 percent
acrylamide gels. The bands of correct size were
25   excised from the gels, and the DNA was eluted from
the gel matrix.

The DNA fragments of (1-4) and (5-6) were
ligated together to form a 177 base pair molecule.
As above, the ligation was digested with restriction
30   enzymes to create the necessary ends for subsequent
cloning of the molecule. In this case, the ligation
of fragments (1-4) and (5-8) was digested with Nco I
and Sph I. The digested ligation product DNA segment
was run on a 6 percent polyacrylamide gel. The band
35   of 177 base pairs was excised and eluted from the
gel.

The 177 base pair fragment was cloned into
plasmid pARC466. Plasmid pARC466 is a plasmid
identical to M13mp19 except that an Nco I site has

-141-

replaced the native Hind III site. This plasmid
contains a polylinker region including a Sma I site
that is downstream from the Sph I site.

    The Nco I site in plasmid pARC466 was created

5   by replacing the originally present Hind III site
using in vitro mutagenesis as discussed previously.
The primer used was:

<u>Nco I</u>
5' CCT GCA GGC ATC CAA CCA TGG CGT

10   (SEQ ID NO:75)
AAT CAT GGT CAT 3'

    Plasmid pARC466 was digested with Nco I and
Sph I. The 177 bp transit peptide DNA fragment ends

15   were designed to clone into these sites. The
ligation of the 177 base pair fragment into plasmid
pARC466 resulted in plasmid pARC480. Plasmid pARC480
was sequenced by M13 protocol to check the sequence
of the designed peptide, which sequence was found to

20   be correct.

ii. Plasmid pATC212

    The transit peptide was moved into a plasmid
that contained a plant promoter and termination

25   sequence. pCaMVCN is a plasmid supplied by Pharmacia
that contains the 35S promoter and a NOS
polyadenylation sequence. The transit peptide was
cloned next to the 35S promoter as follows:

    a) Plasmid pCaMVCN was digested with the

30   restriction enzyme Sal I. Linker #1104 from New
England Biolabs d(TCGACCCGGG) was digested with Sal I
and then ligated with the digested pCaMVCN to create
plasmid pATC209.

    b) Plasmid pATC209 was digested with

35   Sma I. Plasmid pARC480 was digested with Nco I and
Sma I to remove the transit peptide. The Nco I site

of the transit peptide DNA

Klenow fragment of <u>E. coli</u>

blunt end to make that fra

Sma I site of plasmid pATC

5   ended fragment was cloned

plasmid pATC209 to create

iii. Plasmid pATC

Plasmid pATC1616 i:

10   pGA482 that contains the g

dehydrogenase-4H with the

frame with the coding sequ

dehydrogenase-4H gene. Th

by the <u>CaMV 35S</u> promoter a

15   polyadenylation site downs1

gene. The plasmid was mad

The plasmid pATC16(

version of the phytoene del

site at the initiation met

20   pATC1607 was digested with

site is the same as the Nc

in Figure 5 and is the Nco

1510 in Figure 11. The Nc

by treating with the Kleno

25   polymerase.

The thus treated pl

then digested with Sph I.

production of about 1506 b

the structural gene for ph

30   At the 5' end of the fragme

the 3' end of the fragment

Plasmid pATC212 was

Sma I. The Sph I site is a

transit peptide sequence an

35   downstream in the polylinke

-143-

pATC212. The above Sph I-blunt ended phytoene
dehydrogenase-4H gene fragment was cloned into the
pATC212 plasmid, resulting in plasmid pATC1612.

     Plasmid pATC1612 contains the <u>CaMV 35S</u>

5   promoter, the transit peptide sequence, the
structural phytoene dehydrogenase-4H gene, and the
NOS polyadenylation sequence. This whole region of
pATC1612 can be moved as an Xba I-Xba I fragment,
since there are Xba I sites upstream from the <u>CaMV</u>

10   <u>35S</u> promoter and downstream from the NOS
polyadenylation sequence.

     Plasmid pATC1612 was digested with Xba I and
the about 2450 bp Xba I-Xba I fragment (450 bp <u>CaMV</u>
<u>35S</u> promoter, 177 bp transit peptide sequence, 1506

15   bp phytoene dehydrogenase-4H gene, and the 300 bp NOS
polyadenylation sequence) was cloned into the Xba I
site of plasmid pGA482. The resulting plasmid is
pATC1616.

20   **b. Production in the Plant Cytoplasm**

     To prepare lycopene in the cytoplasm, the
carotenoid genes described before are introduced into
appropriate vector(s), as also described above for
chloroplasts, using identical techniques, except that

25   the transit peptide is eliminated. Because they are
not targeted to the chloroplast, the enzymes remain
in the cytoplasm, and, acting on the ubiquitous
isoprenoid intermediate, farnesyl pyrophosphate,
produce lycopene in the cytosol.

30

Example 15. Lycopene Cyclase Gene
      a. Localization

     The location of the lycopene cyclase gene on
pARC376 was established as described before for the

35   other enzyme genes. If the gene for lycopene cyclase

-144-

were deleted, mutated or otherwise impaired, there
would not be an active lycopene cyclase enzyme and
lycopene would accumulate. Lycopene imparts a red
color to <u>E. coli</u> cells producing it, whereas beta-
5   carotene imparts a yellow color to <u>E. coli</u> cells
producing beta-carotene.

The following experiments demonstrated that
the gene is located on a 1548 bp DNA fragment of
plasmid pARC376 bounded by the Sal I site (9340) and
10   the Pst I site (7792) shown in Figure 5.

Plasmid pARC376 was partially digested with
Ava I, the ends were religated, and the plasmid DNA
was transformed into <u>E. coli</u> cells strain HB101.
This plasmid, named pARC376-Ava 102, contained a 611
15   bp Ava I fragment deletion from position 8231 to 8842
and also a 1611 bp Ava I fragment deletion from
position 8842 to 10453.

Some <u>E. coli</u> cells transformed with the Ava I
digested pARC376 plasmid were found to have impaired
20   lycopene cyclase gene function, and therefore,
accumulated lycopene. These results indicated that
the gene for lycopene cyclase was present in the
region near the Sal I site at 9340.

25   **b. Plasmid pARC1009**

Example 8b describes the construction of
plasmid pARC137B, whose <u>Erwinia</u> <u>herbicola</u> DNA insert
is diagrammatically illustrated below.

30 Nco Sal   Hind   Sal   Sal   Hind
    I* I*   III   I   I   III
    |---|------|----------|----------|-------||------------|
              (13463)   (10432)   (9340)

35   The Nco I and Sal I sites in the above
diagram with asterisks are in the polylinker portion
of parent plasmid pARC306A.

-145-

Plasmid pARC137B was digested with Sal I and
then the region from the polylinker Sal I site to the
Sal I site at about original position 9340 was
ligated back together, to form plasmid pARC137-5. A
5   Sal I-Sal I fragment of about 4123 bp was thereby
removed. The formed plasmid retained the Rec 7
promoter that was now adjacent to the Erwinia
herbicola DNA beginning at about the Sal I site at
about original position 9340.
10   The resulting plasmid also contained two Stu
I restriction sites between the remaining Sal I and
Hind III sites. Those Stu I sites were at about
original positions 7306 and 3538.
Digestion of plasmid pARC137-5 with Stu I,
15   and religation of the Stu I-terminated fragments
containing the above-illustrated Nco I and Hind III
sites resulted in a new plasmid named pARC1009. That
plasmid contained Erwinia herbicola DNA of interest
from the Sal I site originally at about position 9340
20   to the Stu I site originally at about position 7306,
and the Rec 7 promoter adjacent to that Sal I site.
Plasmid pARC1009 was transformed into
E. coli, strain JM101, and the cells were grown and
treated with nalidixic acid to induce the Rec 7
25   promoter. The protein fraction was isolated,
analyzed on PAGE and a dominant protein band of 36
kilodaltons was noted. This protein band was
identified as the enzyme lycopene cyclase, as
discussed hereinafter. The protein band was isolated
30   and subjected to N-terminal amino acid sequencing.
The first 25 N-terminal amino acid residues were
determined as shown in Figure 19.
Comparison of the N-terminal amino acid
sequence of the lycopene cyclase enzyme with the DNA
35   sequence of the pARC376 plasmid revealed the position

of the initiation codon of the lycopene cyclase gene.
Surprisingly, the initiation codon is GTG, not the
much more common ATG. A GTG codon normally codes for
the amino acid valine, but under rare instances in
5   bacteria, it can also code for methionine when it is
the first amino acid in a protein (G.D. Stormo, 1986,
in Maximizing Gene Expression, W. Reznikoff, L. Gold
(Eds) Butterworths, Stoneham, Massachusetts, pp 195-
224.) Thus, from this comparison, the 5' end of the
10   gene for lycopene cyclase was found to begin about
338 bp downstream from the Sal I site at original
position 9340.

     c.   Plasmid pARC465
15   A series of studies was performed to
determine the location of the 3' end of the gene. A
plasmid, pARC465, which contains the carotenoid genes
for GGPP synthase, phytoene synthase, phytoene
dehydrogenase-4H and the chloramphenicol
20   acetyltransferase gene that confers resistance to the
antibiotic chloramphenicol, was constructed as
follows.
          The plasmid pARC307D is an analogous plasmid
to the plasmid pUC8, except that plasmid pARC307D
25   contains the chloramphenicol acetyltransferase gene
instead of the ampicillinase gene. Plasmid pARC307D
also contains the same polycloning linker as pUC8.
          Plasmid pARC307D was digested with Hind III
and Eco RI. The plasmid pARC376-Ava 102 (Example 9b)
30   was also digested with Hind III and Eco RI. The
resultant about 8000 bp fragment from Hind III
(13463) to Eco RI (3370) of plasmid pARC376-Ava 102
was isolated from an agarose gel (the fragment size
is only about 8000 bp because the Ava I deletions in
35   plasmid pARC376-Ava 102 described before deleted

about 2200 bp from the parent pARC376 plasmid). This
about 8000 bp Hind III-Eco RI fragment was cloned
into the Hind III- and Eco RI-digested plasmid
pARC307D. The resulting plasmid, pARC465, caused the
5   production of lycopene when transformed into E. coli,
and also conferred resistance to the antibiotic
chloramphenicol.

The plasmid pARC1009, which contains the gene
for lycopene cyclase, was introduced into E. coli
10   cells containing plasmid pARC465, and the cells were
grown on chloramphenicol and ampicillin. These cells
produced beta-carotene. This indicated that the 3'
end of the gene for lycopene cyclase was upstream
from the Stu I site (original position about 7306).

15

d. Plasmid pARC1008

To further define the location of the 3' end
of the gene, the 1548 bp Sal I (9340) to Pst I (7792)
DNA fragment (Example 15a) was cloned into plasmid
20   pARC306A. The resulting plasmid, pARC1008, was
introduced into E. coli cells that already contained
plasmid pARC465. These cells, grown in the presence
of chloramphenicol and ampicillin, produced beta-
carotene. These results indicated that the 3' end of
25   the gene was present upstream from the Pst I (7792)
site.

In summary then, the gene for lycopene
cyclase is contained in an about 1548 bp Sal I to Pst
I fragment of plasmid pARC376. The actual initiation
30   codon is about 338 bp downstream from the Sal I site.
Therefore, the bounds of the gene for lycopene
cyclase are approximately from position 9002 to the
Pst I site at position 7792 in Figure 5, enclosing an
approximately 1210 bp DNA segment. Figure 19
35   contains a nucleotide sequence obtained from single

-148-

strand sequencing and a partial amino acid sequence
obtained by amino acid sequencing for lycopene
cyclase.

5   Several constructs have been made in which
the 5' end of the gene for lycopene cyclase has been
modified. Two are described below.

     e. Plasmid pARC147
          In one construct, the initiation codon was
10   changed from a GTG sequence to an ATG sequence by
introducing a Nco I site by in vitro mutagenesis at
the beginning of the gene as follows. An
oligonucleotide probe was synthesized that had the
following sequence as compared with the normal
15   sequence:

Native DNA Sequence
               *MET   (SEQ ID NO:76)
     A GAG CGT ATC GTG AGG GAT CTG ATT TTA GTC GGC G
20

New DNA Sequence

               Nco I   (SEQ ID NO:77)
     G CGC GGA TCC ATG GGG GAT CTG ATT TTA GTC GGC G
25   *MET

_____

     *   Initiation Methionine; bold-faced letters are as
         described before.
30

          The Nco I restriction site sequence is CC
ATGG, therefore, the new sequence at the initiation
methionine introduced an Nco I site. This newly
modified lycopene cyclase gene, starting at the
35   introduced Nco I site was cloned into the plasmid
pARC306A to generate the plasmid pARC147. Plasmid
pARC147 was introduced into E. coli cells already
containing plasmid pARC465, and the cells were grown

-149-

in the presence of chloramphenicol and ampicillin.
These cells produced beta-carotene. Thus, a
functional lycopene cyclase gene within an about 1210
bp DNA fragment from Nco I to Pst I that can be moved
5   into other plasmids for the expression of the enzyme,
was constructed.

        f. Lycopene Cyclase Assay

                Cultured <u>E. coli</u> cells separately transformed
10   with plasmid pARC1606, described below, that cause
lycopene accumulation in <u>E. coli</u>, and with plasmid
pARC147, discussed before, that contains the <u>Rec 7</u>-
driven lycopene cyclase gene were separately
homogenized. The homogenates were mixed at a ratio
15   of 1:1 in the presence of 2.5 mM $MgCl_2$, 3 mM $MnCl_2$, 4
mM ditheriothreitol, and 6 mM ATP for six hours at
30°C.

                The assay mixture was thereafter lyophilized
and extracted with acetone:methanol (7:2, v:v). The
20   extract was concentrated and analyzed by HPLC.
β-Carotene was detected; about 54 ng of the <u>cis</u>
isomer and about 27 ng of the <u>trans</u> isomer. Thus,
the genetically engineered gene for lycopene cyclase
present in plasmid pARC147, was actively transcribed
25   by the transformed <u>E. coli</u> host cells.

                Cofactors such as FAD, NADP and FMN are not
required for lycopene cyclase activity. ATP is,
however, essential for activity.

30   <u>Construction of pARC1606</u>
                The construction of plasmid pARC1606
proceeded with a series of intermediate vectors.
                The plasmid pARC376 was partially digested
with Bam HI and then religated. The religated
35   plasmid was transformed into <u>E. coli</u> cells and cells

were selected that contained a plasmid in which Bam
HI fragments of about 1045 bp (from original position
3442 to 4487) and of about 815 bp (from original
position 5302 to 4487) were deleted from the pARC376

5   plasmid. The name of the new plasmid was pARC376-Bam
100, and the plasmid caused the E. coli cells to
produce β-carotene, since the gene for β-carotene
hydroxylase was deleted.

The plasmid pARC376-Bam 100 was digested with

10   Hind III and Eco RI. The fragment containing the
Erwinia herbicola carotenoid genes was isolated and
religated. The coordinates for the Hind III and Eco
RI sites originally from plasmid pARC376 are 13463
and 3370, respectively.

15   Plasmid pARC307D, supra, also contains the
pUC8 MCS. Plasmid pARC307D was digested with Hind
III and Eco RI, and the Erwinia herbicola Hind III
and Eco RI fragment excised from plasmid pARC376-Bam
100 was cloned into plasmid pARC307D to form plasmid

20   pARC279. This plasmid conferred chloramphenicol
resistance to the E. Coli cells and also caused them
to produce β-carotene. The plasmid pARC279 contains
about 11.7 kb.

Plasmid pARC279 was partially digested with

25   Bgl II and Bam HI and then religated to delete
specific regions from the pARC279 plasmid that were
not necessary for β-carotene production and make the
plasmid as small as possible. A clone was found in
which the size of the plasmid was about 10 kb (about

30   1.7 kb had been deleted), that conferred
chloramphenicol resistance to E. coli and caused the
synthesis of β-carotene. That plasmid was named
pARC281B.

Plasmid pARC1606 was made from plasmid

35   pARC281B by mutagenizing E. coli cells that contained

-151-

plasmid pARC281B with nitrosoguanidine (NTG)
according to the following protocol.

The following is the NTG mutagenesis
protocol:

5   1. <u>E. coli</u> cells containing plasmid pARC281B
           were grown to log phase - about $3-5 \times 10^8$
           cells/ml or Absorbance 0.3-0.6 at 600 nm.
       2. The cells were washed twice with
           phosphate buffer (50 mM, pH 7.0), and
10   then resuspended in 1/10th of the
           original volume of growth medium.
       3. NTG was added to the cells in phosphate
           buffer to a final concentration of 100
           μg/ml. The cells were incubated for 1
15   hour at 37°C.
       4. The cells were washed three times in
           phosphate buffer to remove the NTG. The
           cells were then resuspended in Luria-
           Broth with 25 μg/ml of chloramphenicol
20   and grown for about 15-18 hours at 37°C.
       5. The cells were then diluted and plated on
           Luria-Broth Medium with 1.5 percent Agar
           with 25 μg/ml chloramphenicol. A colony
           was found that produced lycopene as
25   evidenced by the red appearance of the
           colony. The plasmid contained in that
           colony was isolated and called plasmid
           pARC1606.

       A mutation was induced somewhere in the gene
30 for lycopene cyclase after the nitrosoguanidine
    treatment that caused the inactivation of the enzyme.
    This caused the cells to accumulate lycopene, the
    precursor to β-carotene. Cells that contained the
    plasmid with this mutation were now red, due to the

accumulation of lycopene, instead of the β-carotene
yellow color.

    Cells containing plasmid pARC1606 were used
as a source of lycopene for the lycopene cyclase
5   assays described before.

    g. Plasmid pARC1509

    The new construct, plasmid pARC147, that
works effectively in <u>E. coli</u>, is not effective in
10   yeast. It appears that the second N-terminal amino
acid, which was changed from Arg to Gly by the above
procedure, made this gene inactive in yeast.
Therefore both 5' and 3' ends of the lycopene cyclase
gene were genetically re-engineered as follows.

15   A Sph I restriction site at the initiation
Met codon and a Bam HI restriction site at the 3' end
of the gene were introduced into the native sequence
by PCR (as described before) using the following
probes:

20

    For the Sph I site at the 5' end

       <u>Sph I</u>   (SEQ ID NO:78)
    5' G CGG **CGC ATG** CGG GAT CTG ATT TTA GTC GGC G 3'

25   For the Bam HI site at the 3' end

       <u>Bam HI</u>   (SEQ ID NO:79)
    5' CAT CGG **ATC** CTG TCA GGA AAA TGG TTC AGC 3'

30   An about 3012 bp fragment from Sal I (9407)
to the Nco I site (6395) was excised from the plasmid
pARC271D described in Example 8. This fragment was
used as the template for the PCR reaction that was
performed as described previously.

35   After PCR, the reaction mixture was digested
with Sph I and Bam HI. The about 1142 bp fragment
shown in Figure 19, between the first G residue of

-153-

the Sph I (18) site and the first G residue of the
Bam HI (1168) site, was isolated on agarose gel as
previously described. This about 1142 bp Sph I-Bam
HI fragment of the cyclase gene was cloned into pUC18
5   which had been previously digested with Sph I and Bam
HI. The resulting plasmid was called pARC1509.

       h.   Plasmid pARC1510
              To determine whether the genetically
10   engineered version of the lycopene cyclase gene in
plasmid pARC1509 codes for an active protein, the
structural gene segment was introduced adjacent to
the TAC promoter in the plasmid pKK223-3 (Pharmacia)
as follows. Upstream from the Sph I site of plasmid
15   pARC1509 (in the polycloning sequence) is a unique
Hind III site. The plasmid pARC1509 was digested
with Hind III and Bam HI, an about 1156 bp Hind III-
Bam HI fragment was isolated. The fragment ends were
made blunt by treatment with the Klenow fragment of
20   DNA Polymerase I.
              The plasmid pKK223-3 contains a unique Eco RI
site adjacent to the TAC promoter. Plasmid pKK223-3
was digested with Eco RI and the ends were likewise
blunted with the Klenow reagent. The fragment
25   containing the structural gene segment for lycopene
cyclase was ligated into the blunted Eco RI site
adjacent to the TAC promoter to produce the plasmid
pARC1510.
              To verify that the gene for lycopene cyclase
30   was capable of expressing an active protein, plasmid
pARC1510 was introduced into E. coli cells that
already contained the plasmid pARC465 that contains
the CAT resistance gene and the genes necessary to
produce lycopene, but from which the gene for
35   lycopene cyclase had been deleted. E. coli cells

-154-

containing both plasmids pARC465 and pARC1510, were
grown with both chloramphenicol and ampicillin, and
produced beta-carotene.

5   **Example 16.** **Beta-carotene production in _E. coli_**
         a.   **Method One - Plasmid(s) containing engineered**
               genes for GGPP synthase, phytoene synthase,
               phytoene dehydrogenase-4H and lycopene
               cyclase

10   Four carotenoid enzyme genes are required to
     produce beta-carotene from ubiquitous precursors,
     i.e., the genes for GGPP synthase, phytoene synthase,
     phytoene dehydrogenase-4H, and lycopene cyclase. In
     one example, the first three genes, i.e., for GGPP

15   synthase, phytoene synthase, and phytoene
     dehydrogenase-4H enzymes, were present on the plasmid
     pARC465. This plasmid also contains the
     chloramphenicol acetyltransferase gene that confers
     resistance to the antibiotic chloramphenicol in

20   _E. coli_.
          The plasmid pARC1009, described in Example
     15, contains the about 2038 bp Sal I to Stu I DNA
     fragment inserted into plasmid pARC306A. When
     plasmid pARC1009 was transferred to _E. coli_ cells

25   that contained the plasmid pARC465, the cells
     produced beta-carotene at a level of about 0.05
     percent (dry weight).
          The plasmid pARC147, also described in
     Example 15, contains the about 1215 bp Nco I to Pst I

30   fragment that was inserted into the pARC306A plasmid.
     This plasmid was also introduced into _E. coli_ cells
     that contained the plasmid pARC465, and those cells
     also synthesized beta-carotene at a level of about
     0.05 percent (dry weight). Because it was

35   subsequently discovered that this version of the

lycopene cyclase structural gene was inactive in
yeast, its use was discontinu d and the gene was
altered as described in Example 15 to produce plasmid
pARC1510. Plasmid pARC1510, transferred in
5   combination with plasmid pARC465, produced beta-
carotene in E. coli.

   b.   Alternate Method - Plasmid pARC376 with a
        defective gene for beta-carotene hydroxylase
10   The plasmid pARC376 has a sufficient gene
complement to effectuate the synthesis of carotenoids
up to and including zeaxanthin diglucoside in
E. coli. Beta-carotene is the metabolic substrate
for the beta-carotene hydroxylase enzyme that adds
15   two hydroxyl groups at the 3 and 3' positions of
beta-carotene to produce zeaxanthin. If the gene for
beta-carotene hydroxylase is deleted, mutated, or in
some other way made non-functional, the cells
accumulate the substrate beta-carotene.
20

   i.   Plasmid pARC376-Pst 102
        The gene for beta-carotene hydroxylase is
contained on a 975 bp DNA fragment bounded by a Pst I
site (4886) and the Sma I site (5861) in plasmid
25   pARC376. To delete part of the gene for this enzyme,
plasmid pARC376 was partially digested with Pst I,
and the appropriate cut ends were religated.
Analysis of the plasmid DNA determined that the 392
bp Pst I fragment from original position 4886 to 5215
30   was deleted. This plasmid was named pARC376-Pst 102.
        After transformation of plasmid pARC376-Pst
102 into E. coli, colonies with an orange-yellow
color were picked and analyzed for carotenoid content
by methods described before. The normal color of
35   E. coli colonies containing the intact pARC376

-156-

plasmid and producing zeaxanthin diglucoside is
yellow. Analysis of the orange-yellow colored
colonies revealed that only beta-carotene was being
produced at a level of about 0.1 percent (dry

5   weight).

        ii. Plasmid pARC376-Bam 100
        In an analogous procedure, plasmid pARC376
was partially digested with Bam HI and appropriately

10   religated, causing the deletion of an approximately
815 bp fragment from about original position 4487 to
5302. The resultant plasmid was called pARC376-Bam
100. The plasmid DNA was transformed into E. coli
HB101, and orange-yellow colonies were selected and

15   analyzed for carotenoid content. Beta-carotene
accumulated in these cells at a level of about 0.1
percent.

        Example 17. Production of beta-carotene in S.

20   cerevisiae
        The structural gene for each of the four
enzymes required for beta-carotene synthesis is
placed adjacent to an appropriate promoter and
termination sequence that will properly function in

25   S. cerevisiae. Appropriate promoters include the GAL
1 and GAL 10 divergent promoters, described in the
Detailed Description and Example 5, and the
phosphoglyceric acid kinase gene promoter (PGK),
likewise described. An appropriate terminator is the

30   termination sequence from the PGK gene.
        The structural genes for GGPP synthase and
phytoene synthase are present in the plasmid
pARC145G, adjacent to the GAL 10 and GAL 1 promoters
as described in Example 5. The termination sequence

35   from the PGK gene is at the 3' end of the gene for

-157-

phytoene synthase. To produce beta-carotene it was necessary to introduce the genes for phytoene dehydrogenase-4H and lycopene cyclase in vectors that direct the expression of these genes in this

5   microorganism.

One approach to induce beta-carotene synthesis in yeast is to insert these two genes into a vector, such as pARC146, that contains the GAL 10 and GAL 1 divergent promoters and introduce the

10   resultant plasmid into S. cerevisiae that already contains plasmid pARC145G. The resulting population has all of the genetic material required to produce beta-carotene in a form that permits high level expression of the genes.

15

a.   Plasmid pARC1520

The plasmid pARC146D (Example 10) already contains the gene for phytoene dehydrogenase-4H adjacent to the GAL 1 promoter. The structural gene

20   for lycopene cyclase described in Example 15 was cloned into plasmid pARC146D adjacent to the GAL 10 promoter as follows:

The plasmid pARC1509, described in Example 15, was digested with Hind III and Bam HI. The about

25   1156 bp fragment containing the structural gene for lycopene cyclase was isolated and the ends were blunted by treatment with the Klenow fragment of DNA Polymerase I.

Plasmid pARC146D was digested with Eco RI

30   (restriction site is unique in plasmid pARC146D - see Figure 14). The ends of the Eco RI digested plasmid were also blunted and the lycopene cyclase gene was cloned into plasmid pARC146D to produce the plasmid pARC1520. Plasmid pARC1520, therefore, contains the

35   gene for phytoene dehydrogenase-4H adjacent to the

-158-

GAL 1 promoter, the gene for lycopene cyclase
adjacent to the GAL 10 pr moter, and the URA 3 gene
(described before) useful for selection in yeast.
Plasmid pARC1520 was introduced into the S.
5   cerevisiae, strain yPH499, which already contained
the plasmid pARC145G. Beta-carotene was produced at
the level of about 0.01 percent of the dry weight.

Example 18. Production of beta-carotene in Higher
10   **Plants.**
    a. Chloroplast
        Although beta-carotene is synthesized in the
chloroplasts of plants, most higher plant species do
not accumulate very high levels of it. Carrot roots
15 are among the best accumulators, but even in these
the concentration is only about 0.01-0.1 percent (dry
weight). The objective, then, is to increase the
catalytic activity of lycopene cyclase and thereby
the accumulation of beta-carotene.
20   Lycopene production is thought to be the
divergence point of carotenoid synthesis. In one
branch, lycopene is converted to alpha-carotene that
in turn is converted to lutein. Lutein is the
carotenoid that accumulates to the highest
25 concentration level of all carotenoids. In the other
branch, lycopene is converted to beta-carotene, which
does not accumulate to as high a level as the lutein.
If the level for the enzyme for lycopene cyclase is
increased, however, beta-carotene accumulates to
30 higher levels.
        To increase the level of lycopene cyclase in
the chloroplast, the following steps are taken. The
177 bp Nco I to Sph I DNA fragment for the tobacco
transit peptide sequence of plasmid pATC212 is linked
35 in frame with the lycopene cyclase structural gene

having a Sph I site at the initiation codon for
methionine of plasmid pARC1509. This procedure is
analogous to that described in Example 14, except
that the 3'-Bam HI site of the lycopene cyclase gene,
rather than the 3'-Nco I site, is blunted with the
Klenow fragment of DNA polymerase.

The resulting plasmid is cleaved with Xba I,
and the Xba I-Xba I fragment containing the CaMv 353
promoter (about 450 bp), the transit peptide (about
177 bp), lycopene cyclase (about 1150 bp), and NOS
polyadenylation (about 300 bp) sequences are cloned
into plasmid pGA482. That Xba I-Xba I fragment
contains about 2077 bp. This plasmid is transformed
into <u>Agrobacterium tumefaciens</u>, strain A281.

The relevant features of pGA482 were
described in Example 14 and include (i) the left and
right borders of the T-DNA sequence, which directs
the integration of the DNA sequences between these
borders into the plant genome; (ii) the kanamycin
resistance gene using the NOS promoter for
expression, which allows the selection of kanamycin
resistant plants containing the lycopene cyclase
gene; and (iii) an origin of replication that allows
the replication of pGA482 in <u>Agrobacterium
tumefaciens</u>. pGA482 was introduced into
<u>Agrobacterium tumefaciens</u>, strain A281.

Subsequently, plants, such as tobacco and
alfalfa, are infected with this <u>Agrobacterium</u>. The
gene for lycopene cyclase is expressed under the
influence of the <u>CaMV 35S</u> promoter and is directed to
the chloroplast by the tobacco transit peptide
sequence described before. The resultant plants
produce an increased amount of lycopene cyclase,
which results in a concomitant increased level of
production and accumulation of beta-carotene.

-160-

Other carotenoid enzyme-specific genes can
also be utilized in conjunction with the lycopene
cyclase gene to increase the production and
accumulation of beta-carotene. These include genes

5   for GGPP synthase, phytoene synthase, and phytoene
dehydrogenase-4H. The introduction of these genes
into higher plants involves the same manipulations as
described above for lycopene cyclase. The genes are
attached to the tobacco transit peptide DNA sequence

10   and are then placed adjacent to a functional plant
promoter, such as the CaMV 35S promoter. Also placed
adjacent, is a polyadenylation sequence, such as the
NOS polyadenylation sequence.
            These gene constructs are introduced into

15   plants along with the gene for lycopene cyclase, and
the combination results in increased total enzyme
activity in this portion of the carotenoid synthesis
pathway. This further results in an increase of
beta-carotene synthesis and accumulation in the

20   chloroplast.

        b.   Cytoplasm
            Introducing Erwinia herbicola genes for GGPP
synthase, phytoene synthase, phytoene dehydrogenase-

25   4H, and lycopene cyclase results in beta-carotene
synthesis in the cytoplasm. In order to express
these enzymes in plant cells, the structural genes
are individually cloned into one or more vectors that
contain a promoter and a polyadenylation sequence

30   that will function in plants. One such vector is the
before-described pCaMVCN, with the CaMV 35S promoter
and the NOS polyadenylation sequence. The four genes
with the appropriate promoters and polyadenylation
signals are then inserted into the before-described

35   plasmid, pGA482.

-161-

Plasmid pGA482, containing the four carotenoid-specific genes with the appropriate regulatory signals, is transformed into A. tumefaciens, such as strain A281. Subsequently,

5   plants such as tobacco and alfalfa are infected with the A. tumefaciens, containing the four carotenoid genes, during which process, the carotenoid genes are transfected and integrated into the plant genome. The result is that the transformed plants have the

10   necessary genes, and the capacity to produce and accumulate beta-carotene in the cytoplasm. The CaMV 35S promoter causes the carotenoid genes to be expressed.

15   **Example 19.** *β*-**Carotene Production in Pichia pastoris**
       The before-described method is also extendable to other yeasts. One yeast system that serves as an example is the methylotrophic yeast, Pichia pastoris.

20   To produce *β*-carotene in P. pastoris, structural genes for GGPP synthase, phytoene synthase, phytoene dehydrogenase-4H and lycopene cyclase are placed under the control of regulatory sequences that direct expression of structural genes

25   in Pichia. The resultant expression-competent forms of those genes are introduced into Pichia cells.
       For example, the transformation and expression system described by Cregg et al., Biotechnology 5:479-485 (1987); Molecular and

30   Cellular Biology 12:3376-3385 (1987) can be used. A structural gene for GGPP synthase such as that from plasmid pARC489D is placed downstream from the alcohol oxidase gene (AOX1) promoter and upstream from the transcription terminator sequence of the

35   same AOX1 gene. Similarly, structural genes for

-162-

phytoene synthase, phytoene dehydrogenase-4H, and
lycopene cyclase such as those from plasmids
pARC140N, pARC146D and pARC1509, respectively, are
placed between AOX1 promoters and terminators. All

5   four of these genes and their flanking regulatory
regions are then introduced into a plasmid that
carries both the P. pastoris HIS4 gene and a
P. pastoris ARS sequence (Autonomously Replicating
Sequence), which permit plasmid replication within

10   P. pastoris cells [Cregg et al., Molecular and
Cellular Biology, 12:3376-3385 (1987)].

The vector also contains appropriate portions
of a plasmid such as pBR322 to permit growth of the
plasmid in E. coli cells. The final resultant

15   plasmid carrying GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H and lycopene cyclase genes,
as well as the various additional elements described
above, is illustratively transformed into a his4
mutant of P. pastoris, i.e. cells of a strain lacking

20   a functional histidinol dehydrogenase gene.

After selecting transformant colonies on
media lacking histidine, cells are grown on media
lacking histidine, but containing methanol as
described by Cregg et al., Molecular and Cellular

25   Biology, 12:3376-3385 (1987), to induce the AOX1
promoters. The induced AOX1 promoters cause
expression of the enzymes GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H and lycopene
cyclase and the production of β-carotene in

30   P. pastoris.

The four genes for GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H, and lycopene
cyclase can also be introduced by integrative
transformation, which does not require the use of an

-163-

ARS sequence, as described by Cregg et al., <u>Molecular</u>
<u>and Cellular Biology</u>, 12:3376-3385 (1987).

Example 20. β-Carotene Production in <u>A. nidulans</u>

5   The genes encoding GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H and lycopene
cyclase as discussed before can be used to synthesize
and accumulate β-carotene in fungi such as
<u>Aspergillus nidulans</u>. Genes are transferred to
10   <u>Aspergillus</u> by integration.

For example, the structural gene for GGPP
synthase is introduced into the <u>E. coli</u> plasmid
pBR322. The promoter from a cloned <u>Aspergillus</u> gene
such as <u>argB</u> [Upshall et al., <u>Mol. Gen. Genet</u>.
15   204:349-354 (1986)] is placed into the plasmid
adjacent to the GGPP synthase structural gene. Thus,
the GGPP synthase gene is now under the control of
the <u>Aspergillus</u> <u>argB</u> promoter.

Next, the entire cloned <u>amds</u> gene [Corrick
20   et al., <u>Gene</u> 53:63-71 (1987)] is introduced into the
plasmid. The presence of the <u>amds</u> gene permits
acetamide to be used as a sole carbon or nitrogen
source, thus providing a means for selecting those
<u>Aspergillus</u> cells that have become stably transformed
25   with the amds-containing plasmid.

Thus, the plasmid so prepared contains the
<u>Aspergillus</u> <u>argB</u> promoter fused to the GGPP synthase
gene and the <u>amds</u> gene present for selection of
<u>Aspergillus</u> transformants. <u>Aspergillus</u> is then
30   transformed with this plasmid according to the method
of Ballance et al., <u>Biochem. Biophys. Res. Commun</u>.
112:284-289 (1983).

The GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H and lycopene cyclase
35   structural genes are each similarly introduced into

the E. coli plasmid pBR322. Promoters for the cloned
Aspergillus argB gene [Upshall et al., Mol. Gen.
Genet, 204:349-354 (1986)] are placed immediately
adjacent to those three structural genes. Thus,

5   these structural genes are controlled by the
Aspergillus argB promoters.

       The entire, cloned Aspergillus trpC gene
[Hamer and Timberlake, Mol. Cell. Biol., 7:2352-2359
(1987)] is introduced into the plasmid. The trpC

10   gene permits selection of the integrated plasmid by
virtue of permitting transformed trpC mutant
Aspergillus cells to now grow in the absence of
tryptophan. The Aspergillus strain, already
transformed with the plasmid containing the GGPP

15   synthase gene, is now capable of synthesizing
β-carotene.

Example 21.   Production of Zeaxanthin in E. coli
**   a. Construction of Plasmid pARC404BH**

20   The about 2938 bp fragment from the Eco RV
site at position 4323 to the Stu I site at position
7306 from plasmid pARC376 (Figure 5) was cloned into
the Sma I site of M13mp19 (obtained from BRL). The
resulting plasmid was named pARC404BH-B. This

25   plasmid was used for the introduction of an Nco I
site at the initiation methionine of the β-carotene
hydroxylase enzyme (position 4991 of Figure 5) using
the method described in Ausabel et al. and discussed
in Example 2(f).

30   The oligonucleotide probe used for the in
vitro mutagenesis to introduce the Nco I site was:

5' T TAA ACT ATT TAG TAC CAT GGC GGT GCG CGC TCC TG 3'
                                     (SEQ ID NO:80)

35   Upon introduction of the Nco I site at the β-carotene
hydroxylase initiation methionine, the following

-165-

changes occurred in the nucleotide sequence of the
enzyme.

Original sequence:

                                        (SEQ ID NO:81)
5   CAG GAG CGC GCA CCG CT ATG CTA GTA AAT AGT TTA A...
                                Met Leu Val Asn Ser Leu ...
                                        (SEQ ID NO:82)

New Sequence:

10
                                Nco I   (SEQ ID NO:83)
       CAG GAG CGC GCA CCG CC ATG GTA CTA AAT AGT TTA A...
                                Met Val Leu Asn Ser Leu ...
                                        (SEQ ID NO:84)

15
           The plasmid with the newly introduced Nco
     I site at the initiation methionine residue was named
     plasmid pARC404BH-C.
           Plasmid pARC404BH-C was digested with Nco
20   I and Bam HI. The about 311 bp Nco I to Bam HI
     fragment (original coordinates from plasmid pARC376
     is 4991 for the newly introduced Nco I site and 5302
     for the Bam HI site) was isolated and cloned into the
     Nco I and Bam HI sites of plasmid pARC466 of Example
25   14. This plasmid was called pARC404BH-A.
           Plasmid pARC376 was digested with Bam HI
     and Sma I, and the about 559 bp fragment from Bam HI
     (5302 of Figure 5) to Sma I (5861 of Figure 5) was
     isolated. The plasmid pARC404BH-A was digested with
30   Bam HI and Sma I.
           The about 559 bp Bam HI-Sma I fragment
     isolated from plasmid pARC376 was cloned into the
     Bam HI and Sma I sites of plasmid pARC404BH-A. The
     resulting plasmid was called pARC404BH. This plasmid
35   contains the structural gene for β-carotene
     hydroxylase with the newly introduced Nco I site at
     the beginning of the gene, whose sequence is included
     in Figure 21. The structural gene can be moved as an
     about 870 bp Nco I-Sma I fragment.

-166-

b. E. Coli Production of Zeaxanthin

Plasmid pARC404BH was digested with Nco I and Sma I, and the about 870 bp Nco I-Sma I fragment was isolated. The plasmid pARC306A shown
5   schematically in Figure 6, which contains the Rec 7 promoter, was digested with Nco I and Sma I. The about 870 bp Nco I-Sma I fragment was cloned into the pARC306A plasmid to form to form plasmid pARC406BH. This plasmid contains the structural gene for
10   β-carotene hydroxylase adjacent to the E. coli Rec 7 promoter.

The following study was carried out to show that plasmid pARC406BH encodes a functional β-carotene hydroxylase enzyme. E. coli cells
15   containing the plasmid pARC279 produce β-carotene, but not zeaxanthin, because the gene for β-carotene hydroxylase has been deleted from the plasmid; Example 15(f). Plasmid pARC279 also contains the chloramphenicol acetyltransferase gene that confers
20   resistance to E. coli cells to the antibiotic chloramphenicol. E. coli cells containing the plasmid pARC279 were further transformed with the plasmid pARC406BH, and then the cells were grown in the presence of chloramphenicol and ampicillin.
25   Pigments were analyzed and zeaxanthin was found. This demonstrated that the gene for β-carotene hydroxylase present in plasmid pARC406BH is an active gene that produces a functionally active enzyme.

The structural gene for β-carotene
30   hydroxylase can be moved from plasmid pARC406BH as an about 870 bp Nco I-Sma I fragment. However, there are restriction sites downstream from the Sma I site in the multiple cloning sequence of plasmid pARC306A that also can be used to move the structural gene
35   from plasmid pARC406BH. The structural gene was

-167-

moved as a Nco I-Hind III fragment using one of those
downstream restriction sites in the construction of
the plasmid pARC145H that is described below.

5   **Example 22.   Construction of Plasmid pARC145H and
the use of Plasmid pARC145H for
Zeaxanthin Production in Yeast**
Plasmid pARC145G contains the structural
genes for GGPP synthase and phytoene synthase
10   adjacent to the GAL 1 and GAL 10 promoters (Figure
10). There is a unique Sph I site in plasmid
pARC145G near the PGK terminator. The gene for β-
carotene hydroxylase coupled with the yeast PGK
promoter and the termination sequence from the URA 3
15   gene was cloned into the Sph I site of plasmid
pARC145G. The resulting plasmid, pARC145H, contained
three carotenoid genes with GGPP synthase and
phytoene synthase genes using the Gal 10/1 promoters
and the β-carotene hydroxylase gene using the PGK
20   promoter. This plasmid transformed in yeast cells
along with the plasmid pARC1520 which contains the
genes for phytoene dehydrogenase and lycopene cyclase
caused the transformed S. cerevisiae cells to produce
zeaxanthin through the conversion of β-carotene to
25   zeaxanthin by the β-carotene hydroxylase enzyme.
The plasmid pARC145H was constructed
through a series of intermediate vectors as described
below.

30   **b.   Isolation of the PGK Promoter**
The gene for 3-phosphoglycerate kinase
(PGK) from Saccharomyces cerevisiae was isolated from
a lambda MG14 library, which was provided by
Dr. Maynard Olsen (Washington University, St. Louis,
35   MO). The strategy to isolate the gene used the

-168-

nucleotide sequence of the PGK gene which had been
determined and published [Hitzeman et al., Nucleic
Acids Res., 10(23):7791-7808 (1983)]. An
oligonucleotide probe was constructed for the region
5   surrounding the initiation methionine of the PGK
enzyme. The probe was as follows:
5'ATA AAG ACA TTG TTT TTA GAT CTG TTG TAA 3' (SEQ ID NO: 100)
            Two nucleotides were changed from the
original sequence and those are shown above in bold.
10 With those two changes, a restriction site for Bgl II
(AGATCT) was made.
            The lambda MG14 library was screened by
hybridization with the above oligonucleotide probe.
A lambda clone was found that contained the PGK gene.
15 A 2.6 kb Hind III-Eco RI fragment was excised from
lambda clone and cloned into the M13mp19 DNA which
had been digested with Hind III and Eco RI. The name
of this resulting plasmid was mARC127. The 2.6 kb
Hind III to Eco RI fragment contains all of the PGK
20 promoter and part of the structural portion of the
PGK gene. (This construct is illustrated in the
restriction map on page 7795 of the above Hitzeman et
al. article.)

25   c. Introduction of Bgl II and Eco RI sites in
            the PGK Promoter
            To make a version of the PGK promoter that
could be used to express heterologous structural
genes, two restriction sites were introduced into the
30 PGK promoter region. These introduced restriction
sites enabled the PGK promoter to be moved as a Eco
RI and Bgl II fragment.
            The Bgl II restriction site was introduced
12 bp upstream from the initiation methionine of the
35 PGK gene. The in vitro mutagenesis protocol of

-169-

Ausabel et al. discussed previously was used to
introduce the Bgl II site in the <u>PGK</u> promoter. The
plasmid mARC127 was used as the starting DNA source
for the modification. The same oligonucleotide probe

5 as listed above for the isolation of the <u>PGK</u> gene was
used to introduce a Bgl II site into the <u>PGK</u>
promoter. The nucleotide sequence of the native <u>PGK</u>
sequence is as follows:

    5' TTA CAA CAA ATA TAA AAA CAA TGT CTT TAT ...

10   <u>MET</u> (SEQ ID NO:85)

Following the <u>in vitro</u> mutagenesis the nucleotide
sequence became

                 <u>Bgl II</u>

15   5' TTA CAA CAG ATC TAA AAA CAA TGT CTT TAT ...
                           <u>MET</u>   (SEQ ID NO:86)

in which bold letters indicate the changed bases.
The plasmid with the Bgl II restriction site

20 introduced into the <u>PGK</u> promoter was named mARC128.
        The restriction site for Eco RI was then
introduced into the mARC128 plasmid 530 nucleotides
upstream from the Bgl II restriction site using the
same <u>in vitro</u> mutagenesis protocol as above. The

25 oligonucleotide probe used for the <u>in vitro</u>
mutagenesis protocol was:

     5' CTT TAT GAG GGT AAC ATG AAT TCA AGA AGG 3'
                               (SEQ ID NO:87)

with bold letters as before. The original nucleotide

30 sequence was:

     5' CCT TCT TGA ATT GAT GTT ACC CTC ATA AAG 3'
                               (SEQ ID NO:88)

The nucleotide sequence with the Eco RI site became:

           <u>Eco RI</u>   (SEQ ID NO:89)

35   5' CCT TCT TGA AAT CAT GTT ACC CTC ATA AAG 3'

        The plasmid with the Eco RI and the Bgl II
restriction sites was named pARC306M. With this

-170-

plasmid, the <u>PGK</u> promoter can be moved as an about
530 bp Eco RI-Bgl II fragment.

d.   Construction of Plasmid pARC135A

5   Plasmid pUC18 was digested with Sma I,
leaving blunted ends. Nco I linkers (CCCATGGG,
obtained from New England Biolabs) were ligated to
the Sma I site of the digested pUC18 plasmid, and the
plasmid was recyclized. The resulting plasmid
10   pSOC109 contained an Nco I site where the Sma I site
was originally.

        Plasmid pSOC109 was digested with Eco RI
and Nco I. The plasmid pARC306M was digested with
Bgl II, the ends were polished using the Klenow
15   fragment of DNA Polymerase to form a blunt end, and
then an Nco I linker was ligated to the blunted Bgl
II position, as before, and the plasmid was
recyclized.

        The treated, recyclized plasmid pARC306M
20   was then digested with Eco RI and Nco I. The about
530 bp <u>PGK</u> promoter with Eco RI and Nco I ends was
isolated from an agarose gel. This <u>PKG</u> promoter-
containing fragment was then cloned into the Eco RI
and Nco I digested pSOC109. The resulting plasmid
25   was called pARC135A.

e.   Construction of Plasmid pARC300T

        A 67 bp oligonucleotide was chemically
synthesized for the <u>URA 3</u> gene terminator. A Hind
30   III site was placed at the 5' end of the terminator
and a Kpn I site was placed at the 3' end of the
terminator. The sequence of this Hind III-Kpn I
fragment is shown below, with underlined restriction
sites:

35

Hind III
5'<u>AGCTT</u>CGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACGCATATT
     AGCTTCTTGCTTCCTTCCTGGTGTCTGAATCTAACCATATATATGCGTATAA

<div align="center">

Kpn I
GCGGCCGC<u>GGTAC</u> 3'   (SEQ ID NO:90)
CGCCGGCGC   (SEQ ID NO:91)
</div>

This Hind III-Kpn I fragment was cloned into the Hind III and Kpn I sites of the plasmid pARC300E to form the plasmid pARC300M. The plasmid pARC300E is a derivative of the plasmid pUC8 with a unique Sma I site between the Aat II and Cla I sites of the plasmid. The restriction map for restriction sites that are present only once in pARC300E is shown in Figure 22, whereas a similar map for plasmid pARC300M is illustrated in Figure 23.

A <u>LEU 2</u> gene not relevant for the present construct was cloned as a blunted fragment into the Sma I site of plasmid pARC300M to form the plasmid pARC300T. A restriction map for plasmid pARC300T similar to those of Figures 22 and 23 is shown in Figure 24.

f.   Construction of Plasmid pARC426BH

The plasmid pARC300T was digested with Eco RI and Hind III. The plasmid pARC135A was digested with Eco RI and Nco I, and the about 530 bp <u>PGK</u> promoter fragment was isolated. The plasmid pARC406BH was digested with Nco I and Hind III and the Nco I-Hind III fragment containing the structural gene for β-carotene hydroxylase was isolated. A tri-ligation was performed with the <u>PGK</u> promoter, the β-carotene hydroxylase structural gene, and the plasmid pARC307T. Graphically this is shown below:

```
Eco RI   Eco RI          Nco I Nco I     Hind III Hind III       Kpn I
----|    |------------|    |------------|    |------------|---
    |    |            |    |            |    |            |
pARC300T PGK Promoter       β-carotene        URA 3 Terminator
                            hydroxylase       pARC300T
```

-172-

Following the tri-ligation, the resulting
plasmid pARC426BH contained the PGK promoter driving the
β-carotene hydroxylase gene with the URA 3 terminator at
the 3' end of the gene. This cassette could be moved as
5   an about 1500 bp Eco RI-Kpn I fragment into other yeast
vectors for expression of this gene in yeast.

   g.   Construction of Plasmid pARC145H

        Plasmid pARC426BH was digested with Eco RI and
10   Kpn I and the about 1500 bp fragment was isolated (about
530 bp for the PGK promoter, about 900 bp for the
β-carotene hydroxylase gene, and about 67 bp for the
URA 3 terminator). The ends were made blunt by
treatment with the Klenow fragment of DNA Polymerase.
15   Sph I linkers (GGCATCC, New England Biolabs) were
ligated to the about 1500 bp fragment. The fragment was
then digested with Sph I. This Sph I-digested fragment
was then cloned into the unique Sph I site of plasmid
pARC145G (Figure 10), which contains the genes for GGPP
20   synthase and phytoene synthase to form the plasmid
pARC145H.

   h.   Production of Zeaxanthin in Saccharomyces
        Cerevisiae

25   Two plasmids were introduced into the S.
Cerevisiae yeast strain YPH499: the plasmid pARC145H,
which contains the genes for GGPP synthase, phytoene
synthase, and β-carotene hydroxylase; and the plasmid
pARC1520, which contains the genes for phytoene
30   dehydrogenase-4H and lycopene cyclase. Doubly
transformed yeast cells were grown in the presence of
galactose to induce the Gal 10 and Gal 1 promoters.
        Under these conditions the transformed yeast
cells produced zeaxanthin at about 0.01 percent of the
35   dry cell weight. The gene for β-carotene hydroxylase

-173-

was expressed using the PGK promoter and the enzyme was
able to convert β-carotene into zeaxanthin. This study
therefore demonstrated that zeaxanthin could be produced
in yeast.

5

Example 23. Zeaxanthin Production in Pichia pastoris

        The before-described method is also extendable
to other yeasts. One yeast system that serves as an
example is the methylotrophic yeast, Pichia pastoris.

10   To produce zeaxanthin in P. pastoris,
structural genes for GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H, lycopene cyclase and beta-
carotene hydroxylase are placed under the control of
regulatory sequences that direct expression of

15   structural genes in Pichia. The resultant expression-
competent forms of those genes are introduced into
Pichia cells.

        For example, the transformation and expression
system described by Cregg et al., Biotechnology 5:479-

20   485 (1987); Molecular and Cellular Biology 12:3376-3385
(1987) can be used. A structural gene for GGPP synthase
such as that from plasmid pARC489D is placed downstream
from the alcohol oxidase gene (AOX1) promoter and
upstream from the transcription terminator sequence of

25   the same AOX1 gene. Similarly, structural genes for
phytoene synthase, phytoene dehydrogenase-4H, and
lycopene cyclase and beta-carotene hydroxylase such as
those from plasmids pARC140N, pARC146D, pARC1509,
pARC145H and pARC406BH, respectively, are placed between

30   AOX1 promoters and terminators. All five of these genes
and their flanking regulatory regions are then
introduced into a plasmid that carries both the
P. pastoris HIS4 gene and a P. pastoris ARS sequence
(Autonomously Replicating Sequence), which permit

35   plasmid replication within P. pastoris cells [Cregg et

-174-

al., <u>Molecular and Cellular Biology</u>, 12:3376-3385
(1987)].

5   The vector also contains appropriate portions
of a plasmid such as pBR322 to permit growth of the
plasmid in <u>E. coli</u> cells. The final resultant plasmid
carrying GGPP synthase, phytoene synthase, phytoene
dehydrogenase-4H, lycopene cyclase and beta-carotene
hydroxylase genes, as well as the various additional
elements described above, is illustratively transformed
10   into a <u>his4</u> mutant of <u>P. pastoris</u>, i.e. cells of a
strain lacking a functional histidinol dehydrogenase
gene.

           After selecting transformant colonies on media
lacking histidine, cells are grown on media lacking
15   histidine, but containing methanol as described by Cregg
et al., <u>Molecular and Cellular Biology</u>, 12:3376-3385
(1987), to induce the <u>AOX1</u> promoters. The induced <u>AOX1</u>
promoters cause expression of the enzymes GGPP synthase,
phytoene synthase, phytoene dehydrogenase-4H, lycopene
20   cyclase and beta-carotene hydroxylase and the production
of zeaxanthin in <u>P. pastoris</u>.

           The five genes for GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H, lycopene cyclase
and beta-carotene hydroxylase can also be introduced by
25   integrative transformation, which does not require the
use of an ARS sequence, as described by Cregg et al.,
<u>Molecular and Cellular Biology</u>, 12:3376-3385 (1987).

Example 24. Zeaxanthin Production in <u>A. nidulans</u>
30   The genes encoding GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H, lycopene cyclase
and beta-carotene hydroxylase as discussed before can be
used to synthesize and accumulate zeaxanthin in fungi
such as <u>Aspergillus nidulans</u>. Genes are transferred to
35   <u>Aspergillus</u> by integration.

-175-

For example, the structural gene for GGPP
synthase is introduced into the E. coli plasmid pBR322.
The promoter from a cloned Aspergillus gene such as argB
[Upshall et al., Mol. Gen. Genet. 204:349-354 (1986)] is
placed into the plasmid adjacent to the GGPP synthase
structural gene. Thus, the GGPP synthase gene is now
under the control of the Aspergillus argB promoter.

Next, the entire cloned amds gene [Corrick
et al., Gene 53:63-71 (1987)] is introduced into the
plasmid. The presence of the amds gene permits
acetamide to be used as a sole carbon or nitrogen
source, thus providing a means for selecting those
Aspergillus cells that have become stably transformed
with the amds-containing plasmid.

Thus, the plasmid so prepared contains the
Aspergillus argB promoter fused to the GGPP synthase
gene and the amds gene present for selection of
Aspergillus transformants. Aspergillus is then
transformed with this plasmid according to the method of
Ballance et al., Biochem. Biophys. Res. Commun. 112:284-
289 (1983).

The GGPP synthase, phytoene synthase, phytoene
dehydrogenase-4H, lycopene cyclase and beta-carotene
hydroxylase structural genes are each similarly
introduced into the E. coli plasmid pBR322. Promoters
for the cloned Aspergillus argB gene [Upshall et al.,
Mol. Gen. Genet, 204:349-354 (1986)] are placed
immediately adjacent to those five structural genes.
Thus, these structural genes are controlled by the
Aspergillus argB promoters.

The entire, cloned Aspergillus trpC gene
[Hamer and Timberlake, Mol. Cell. Biol., 7:2352-2359
(1987)] is introduced into the plasmid. The trpC gene
permits selection of the integrated plasmid by virtue of
permitting transformed trpC mutant Aspergillus cells to

                                    -176-

now grow in the absence of tryptophan. The <u>Aspergillus</u>
strain, already transformed with the plasmid containing
the GGPP synthase gene, is now capable of synthesizing
zeaxanthin.

5

Example 25. Beta-Carotene Hydroxylase in Higher Plants
                Higher plants have the genes encoding the
enzymes required for carotenoid production and so
inherently have the ability to produce carotenoids.
10   Zeaxanthin normally is not accumulated, however, because
zeaxanthin so produced in most plants is further
converted to other products. The carotenoid-specific
genes from <u>Erwinia</u> <u>herbicola</u> described can be used to
express beta-carotene hydroxylase for use by the plant
15   as well as to improve accumulation of zeaxanthin in
plants. Two useful approaches are described below.

        a.   Transport to the Chloroplast
                In the first approach, the gene for beta-
20   carotene hydroxylase of Figure 21 is modified to
introduce the restriction site Sph I in place of the Nco
I site shown in Figure 21 at the initiation methionine
codon, using the method of Ausabel et al. as discussed
before. This change of the Nco I to a Sph I site
25   changes the second amino acid residue in the expressed
enzyme variant to be a leucine residue, so that the
first three residues of the enzyme have the sequence
Met-Leu-Leu.
                Following the procedure of Example 14, the
30   plasmid containing the engineered Sph I site is cleaved
with Sma I and the Sph I to yield a fragment of about
870 bp containing the beta-carotene hydroxylase
structural gene. Plasmid pATC212 that contains the <u>CaMV</u>
<u>35S</u> promoter, the transit peptide and NOS site is
35   cleaved with Sph I and Sma I and the above about 870 bp

-177-

fragment is cloned into the respective sites to form a
plasmid analogous to plasmid pATC1612.

5   Cleavage of the above plasmid with Xba I
provides an Xba I-Xba I fragment that includes about
1797 bp. The Xba I-Xba I fragment contains the
following: the about 450 bp <u>CaMV 35S</u> promoter, (b) the
177 bp transit peptide sequence, (c) the about 870 bp
beta-carotene hydroxylase gene, and (d) the about 300 bp
NOS polyadenylation sequence.

10   This Xba I-Xba I gene construct is inserted
into the plasmid pGA482 (Pharmacia) in a Xba I
restriction site within the multiple cloning linker
region to form a plasmid. The relevant features of
pGA482 include (i) an origin of replication that permits

15   maintenance of the plasmid in <u>Agrobacterium tumefaciens</u>,
(ii) the left and right border sequences from the T-DNA
region that direct the integration of the DNA segment
between the borders into the plant genome, and (iii)
the NOS promoter adjacent to the kanamycin resistance

20   gene that permits plant cells to survive in the presence
of kanamycin.

           The above-formed plasmid is transformed into
<u>Agrobacterium tumefaciens</u> LBA4404 (Clontech, Inc.)
according to standard protocols. <u>Agrobacterium</u> cells

25   containing the plasmid with the transit peptide-beta-
carotene hydroxylase gene construct are transferred by
infection of tobacco leaf discs using the method of
Horsch et al., <u>Science</u>, <u>227</u>:1229-1231 (1985). During
the infection process, the entire DNA segment between

30   the left and right borders of the original pGA482
plasmid is transfected into the plant cells.
Transfected plant cells are selected for kanamycin
resistance, grown under usual conditions and accumulate
zeaxanthin.

A transformed plant as described in Example 18a that contains at least the structural gene for lycopene cyclase, and preferably also contains the genes described herein for GGPP synthase, phytoene synthase and phytoene dehydrogenase-4H, is also useful as a host for transformation as described in this example.

b. Production in the Plant Cytoplasm

The carotenoid genes described before are introduced into appropriate vector(s), as also described above for chloroplasts, using similar techniques, except that the transit peptide is eliminated, and the 5' Nco I site shown in Figure 21 is retained. A plasmid containing the beta-carotene hydroxylase gene such as pARC406BH is cleaved with Nco I. The resulting sticky end is made blunt by filling in with the Klenow fragment of DNA polymerase. Cleavage with Sma I provides a double blunt-ended fragment that can be cloned into the Sma I site of plasmid pATC209 to form a plasmid that contains an about 1620 bp fragment excisable with Xba I that can be cloned into a plasmid such as pGA482 and then used to transform A. tumefaciens.

The resulting, transformed A. tumefaciens is then utilized to transform a higher plant such as tobacco or alfalfa such as that produced in Example 18b. The resulting plants have the necessary genes, and the capacity to produce and accumulate zeaxanthin in the cytoplasm due to expression via the CaMV 35S promoter.

Example 26. Production of Zeaxanthin Diglucoside in E. coli

Example 1 illustrated the production of zeaxanthin diglucoside in E. coli transformed with plasmid pARC376. The discussion below describes the

-179-

production of zeaxanthin diglucoside using <u>E. coli</u>
transformed with two plasmids.

      The location of the DNA sequence that encodes
zeaxanthin glycosylase in plasmid pARC376 was readily
accomplished in view of the work described in the
previous examples related to the preceding five
structural genes. The glycosylase-encoding structural
gene was found to be located just downstream from the
Eco RV restriction site at about position 10256 (Figure
5). The coding sequence was found to begin with the ATG
codon located at position 10232.

      PCR primers designed to yield an Nde I site at
the 5' end and an Ava I site at the 3' end of the
structural gene were prepared. Only three bases in the
native sequence had to be changed. The sequences of the
PCR primers were as follows:

<u>For the 5' end</u>

Native 5'   ATACGCCATGAGCCATTTTGCCATTGTGGC
                                   (SEQ ID NO:92)

                                                   3'
Primer 5'   ATAC<u>CATAT</u>GAGCCATTTTGCCATTGTGGC

                  <u>Nde I</u>   (SEQ ID NO:93)

<u>For the 3' end</u>

Native 5'   GCCCCCCGGGAGTCAGATCGTCTTCATGGA
                                     (SEQ ID NO:94)

                                                   3'
Primer 5'   GCCC<u>CCCGG</u>GAGTCAGATCGTCTTCATGGA

                  <u>Ava I</u>   (SEQ ID NO:95)

                                                   3'

      The template utilized for this mutagenesis
reaction was obtained from a Bam HI-Bam HI (position
10524-7775 of Figure 5) fragment obtained from plasmid
pARC137B (Example 8b). The PCR techniques utilized in
Example 8 were utilized here also.

-180-

The PCR product was digested with Nde I and
Ava I to provide an approximately 1390 bp fragment that
was cloned into the Rec 7 promoter plasmid pARC305N
(Example 3) to form plasmid pARC2019. The Ava I site
5   occurs at position 8842 of plasmid pARC376. Clones were
verified by regenerating the fragment plus vector
pattern on an agarose gel after digestion with Nde I and
Ava I.

E. coli (HB101) that had been transformed with
10   plasmid pARC288 (Example 1b) and produce zeaxanthin were
further transformed with plasmid pARC2019. The
resulting cells were lysed and the cell extracts
analyzed by thin layer chromatography. Extracts from E.
coli cells transformed with plasmids pARC288 and pARC376
15   were used as standards. Zeaxanthin diglucoside was
identified in the extracts from E. coli transformed with
both of plasmids pARC288 and pARC2019.

__Example 27__.   __Production of Zeaxanthin Diglucoside in__
20   __S. cerevisiae__

Zeaxanthin diglucoside is produced in the
yeast S. cerevisiae using the multiply transformed
strain YPH499 prepared in Example 22h that is further
transformed with an appropriate vector that can express
25   zeaxanthin glycosylase in yeast. The preparation of
such a vector is discussed below.

Plasmid pARC2019 is digested with Nde I and
the ends are filled in with the Klenow fragment of DNA
polymerase 1 to form DNA having blunt ends. The
30   synthetic linker

                    GAATTC
                    CTTAAG

is ligated between the two blunt ends.

The resulting circular plasmid is then cleaved
35   with Ava I. A synthetic linker of the sequence

-181-

CCGGGAATTC   (SEQ ID NO:96)
TTAAC

is linked into the cleaved plasmid, and the plasmid is
recircularized.

The above procedures provide a plasmid having
the structural gene for zeaxanthin diglucoside between
two Eco RI restriction sites. That plasmid is then
cleaved with Eco RI to provide an approximately 1401 bp
Eco RI-Eco RI fragment.

Plasmid pSOC713 is cleaved with Eco RI, and
the above approximately 1401 bp Eco RI-Eco RI fragment
is excised and then ligated therein to form a plasmid
appropriate for expression of the zeaxanthin glycosylase
gene in S. cerevisiae, in which the structural gene
encoding zeaxanthin glycosylase is under the control of
the GAL 10 promoter. Transformation of the transformed,
zeaxanthin-producing S. cerevisiae of Example 22h
provides yeasts that produce zeaxanthin diglucoside.

Example 28:   **Zeaxanthin Diglucoside Production in**
                 Pichia pastoris

The before-described method is also extendable
to other yeasts. One yeast system that serves as an
example is the methylotrophic yeast, Pichia pastoris.

To produce zeaxanthin diglucoside in P.
pastoris, structural genes for GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H, lycopene cyclase,
beta-carotene hydroxylase and zeaxanthin glycosylase are
placed under the control of regulatory sequences that
direct expression of structural genes in Pichia. The
resultant expression-competent forms of those genes are
introduced into Pichia cells.

For example, the transformation and expression
system described by Cregg et al., Biotechnology 5:479-
485 (1987); Molecular and Cellular Biology 12:3376-3385
(1987) can be used. A structural gene for GGPP synthase

-182-

such as that from plasmid pARC489D is placed downstream
from the alcohol oxidase gene (AOX1) promoter and
upstream from the transcription terminator sequence of
the same AOX1 gene. Similarly, structural genes for
5   phytoene synthase, phytoene dehydrogenase-4H, lycopene
cyclase, beta-carotene hydroxylase, and zeaxanthin
glycosylase, such as those from plasmids pARC140N,
pARC146D, pARC1509, pARC145H, pARC4068H and pARC2019,
respectively, are placed between AOX1 promoters and
10   terminators. All six of these genes and their flanking
regulatory regions are then introduced into a plasmid
that carries both the P. pastoris HIS4 gene and a
P. pastoris ARS sequence (Autonomously Replicating
Sequence), which permit plasmid replication within
15   P. pastoris cells [Cregg et al., Molecular and Cellular
Biology, 12:3376-3385 (1987)].

        The vector also contains appropriate portions
of a plasmid such as pBR322 to permit growth of the
plasmid in E. coli cells. The final resultant plasmid
20   carrying GGPP synthase, phytoene synthase, phytoene
dehydrogenase-4H, lycopene cyclase, beta-carotene
hydroxylase and zeaxanthin glycosylase genes, as well as
the various additional elements described above, is
illustratively transformed into a his4 mutant of
25   P. pastoris, i.e. cells of a strain lacking a functional
histidinol dehydrogenase gene.

        After selecting transformant colonies on media
lacking histidine, cells are grown on media lacking
histidine, but containing methanol as described by Cregg
30   et al., Molecular and Cellular Biology, 12:3376-3385
(1987), to induce the AOX1 promoters. The induced AOX1
promoters cause expression of the enzymes GGPP synthase,
phytoene synthase, phytoene dehydrogenase-4H, lycopene
cyclase, beta-carotene hydroxylase and zeaxanthin

glycosylase and the production of zeaxanthin diglucoside
in P. pastoris.

     The six genes for GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H, lycopene cyclase,
beta-carotene hydroxylase and zeaxanthin glycosylase can
also be introduced by integrative transformation, which
does not require the use of an ARS sequence, as
described by Cregg et al., Molecular and Cellular
Biology, 12:3376-3385 (1987).

Example 29.   Zeaxanthin Diglucoside Production in
A. nidulans

     The genes encoding GGPP synthase, phytoene
synthase, phytoene dehydrogenase-4H, lycopene cyclase,
beta-carotene hydroxylase, and zeaxanthin glycosylase as
discussed before can be used to synthesize and
accumulate zeaxanthin diglucoside in fungi such as
Aspergillus nidulans. Genes are transferred to
Aspergillus by integration.

     For example, the structural gene for GGPP
synthase is introduced into the E. coli plasmid pBR322.
The promoter from a cloned Aspergillus gene such as argB
[Upshall et al., Mol. Gen. Genet. 204:349-354 (1986)] is
placed into the plasmid adjacent to the GGPP synthase
structural gene. Thus, the GGPP synthase gene is now
under the control of the Aspergillus argB promoter.

     Next, the entire cloned amds gene [Corrick
et al., Gene 53:63-71 (1987)] is introduced into the
plasmid. The presence of the amds gene permits
acetamide to be used as a sole carbon or nitrogen
source, thus providing a means for selecting those
Aspergillus cells that have become stably transformed
with the amds-containing plasmid.

     Thus, the plasmid so prepared contains the
Aspergillus argB promoter fused to the GGPP synthase

-184-

gene and the amds gene present for selection of
Aspergillus transformants. Aspergillus is then
transformed with this plasmid according to the method of
Ballance et al., Biochem. Biophys. Res. Commun. 112:284-
5   289 (1983).

                    The·GGPP synthase, phytoene synthase, phytoene
dehydrogenase-4H, lycopene cyclase, beta-carotene
hydroxylase and zeaxanthin glycosylase structural genes
are each similarly introduced into the E. coli plasmid
10   pBR322. Promoters for the cloned Aspergillus argB gene
[Upshall et al., Mol. Gen. Genet, 204:349-354 (1986)]
are placed immediately adjacent to those six structural
genes. Thus, these structural genes are controlled by
the Aspergillus argB promoters.

15   The entire, cloned Aspergillus trpC gene
[Hamer and Timberlake, Mol. Cell. Biol., 7:2352-2359
(1987)] is introduced into the plasmid. The trpC gene
permits selection of the integrated plasmid by virtue of
permitting transformed trpC mutant Aspergillus cells to
20   now grow in the absence of tryptophan. The Aspergillus
strain, already transformed with the plasmid containing
the GGPP synthase gene, is now capable of synthesizing
zeaxanthin diglucoside.

25   **Example 30. Zeaxanthin Glycosylase and Zeaxanthin
                    Diglucoside in Higher Plants**
                    Higher plants such as tobacco and alfalfa have
the genes encoding the enzymes required for carotenoid
production and so inherently have the ability to produce
30   carotenoids. Zeaxanthin diglucoside normally is not
accumulated, however, because zeaxanthin diglucoside so
produced in most plants is further converted to other
products. The carotenoid-specific genes from Erwinia
herbicola described can be used to express zeaxanthin
35   glycosylase for use by the plant as well as to improve

-185-

accumulation of zeaxanthin diglucoside in plants. Two
useful approaches are described below.

a.   Transport to the Chloroplast

5   In the first approach, the structural gene for
zeaxanthin glycosylase of Figure 25 is modified to
introduce the restriction site Sph I instead of the Nde
I site shown in Figure 25 at the initiation methionine
codon, using a method of Ausabel et al. as discussed

10   before. This process, using plasmid pARC2019, for
example, changes the second amino acid residue of the
enzyme from a serine to an arginine, but provides a
biologically active zeaxanthin glycosylase enzyme
variant.

15   Following the procedures of Example 14, the
plasmid containing the engineered Sph I site is cleaved
first with Ava I and made blunt-ended with the Kenow
fragment of DNA polymerase. Cleavage thereafter with
Sph I provides an about 1200 bp fragment containing the

20   zeaxanthin glycosylase structural gene. Plasmid pATC212
that contains the CaMV 35S promoter, the transit peptide
and NOS polyadenylation site is cleaved with Sph I and
Sma I, and the above about 1200 bp fragment is cloned
into those respective sites to form a plasmid analogous

25   to plasmid pATC1612.
                Cleavage of the above plasmid with Xba I
provides an Xba I-Xba I fragment that includes about
2127 bp. That Xba I-Xba I fragment includes the
following: (a) the about 450 bp CaMV 35S promoter, (b)

30   the 177 bp transit peptide sequence, (c) the about 1200
bp zeaxanthin glycosylase gene, and (d) the about 300 bp
NOS polyadenylation sequence.
                This Xba I-Xba I gene construct is inserted
into the plasmid pGA482 (Pharmacia) in a convenient

35   restriction site within the multiple cloning linker

-186-

region to form a plasmid. The relevant features of
pGA482 include (i) an origin of replication that permits
maintenance of the plasmid in Agrobacterium tumefaciens,
(ii) the left and right border sequences from the T-DNA
5   region that direct the integration of the DNA segment
between the borders into the plant genome, and (iii) the
NOS promoter adjacent to the kanamycin resistance gene
that permits plant cells to survive in the presence of
kanamycin.

10   The above-formed plasmid is transformed into
Agrobacterium tumefaciens LBA4404 (Clontech, Inc.)
according to standard protocols. Agrobacterium cells
containing the plasmid with the transit peptide-
zeaxanthin glycosylase gene construct are transferred by
15   infection of tobacco leaf discs using the method of
Horsch et al., Science, 227:1229-1231 (1985). During
the infection process, the entire DNA segment between
the left and right borders of the original pGA482
plasmid is transfected into the plant cells.
20   Transfected plant cells are selected for kanamycin
resistance, grown under usual conditions and accumulate
zeaxanthin.

               A transformed plant as described in Example
25a that contains at least the structural gene for
25   lycopene cyclase, and preferably also contains the genes
described herein for GGPP synthase, phytoene synthase,
phytoene dehydrogenase-4H and β-carotene hydroxylase is
also useful as a host for transformation as described
herein.

30

       b. **Production in the Plant Cytoplasm**
               The carotenoid genes described before are
introduced into appropriate vector(s), as also described
above for chloroplasts, using similar techniques, except
35   that the transit peptide is eliminated, and the 5' Nde I

site shown in Figure 25 is retained. A plasmid
containing the zeaxanthin glycosylase structural gene
such as pARC2019 is cleaved with Nde I and Ava I. The
resulting sticky ends are made blunt with the Klenow

5　　　　fragment of DNA polymerase. The resulting double blunt-
ended is cloned into the Sma I site of plasmid pARC209
to form a plasmid that contains an Xba I-excisable
fragment that includes the CaMV 35S promoter, zeaxanthin
glycosylase gene and NOS polyadenylation sequence that

10　　　can be cloned into a plasmid such as pGA482 that is then
used to transform A. tumefaciens.

　　　　　　The resulting, transformed A. tumefaciens is
then utilized to transform a higher plant such as
tobacco or alfalfa such as that produced in Example 25b.

15　　　The resulting plants have the necessary genes, and the
capacity to produce and accumulate zeaxanthin
diglucoside in the cytoplasm due to expression via the
CaMV 35S promoter.

　　　　　　Although the present invention has now been

20　　　described in terms of certain preferred embodiments, and
exemplified with respect thereto, one skilled in the art
will readily appreciate that various modifications,
changes, omissions and substitutions may be made without
departing from the spirit thereof. It is intended,

25　　　therefore, that the present invention be limited solely
by the scope of the following claims.

-188-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Ausich, Rodney L
               Brinkhaus, Friedhelm L
               Mukharji, Indrani
               Proffitt, John H
               Yarger, James G
               Yen, Huei-Che B

   (ii) TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
                          Glycosylated Zeaxanthin in
                          Genetically Engineered Hosts

  (iii) NUMBER OF SEQUENCES: 100

   (iv) CORRESPONDENCE ADDRESS:
       (A) ADDRESSEE: Amoco Corp., Patents and Licensing Dept
       (B) STREET: 200 E Randolph St
       (C) CITY: Chicago
       (D) STATE: IL
       (E) COUNTRY: USA
       (F) ZIP: 60680-0703

    (v) COMPUTER READABLE FORM:
       (A) MEDIUM TYPE: Floppy disk
       (B) COMPUTER: IBM PC compatible
       (C) OPERATING SYSTEM: PC-DOS/MS-DOS
       (D) SOFTWARE: PatentIn Release #1.24

   (vi) CURRENT APPLICATION DATA:
       (A) APPLICATION NUMBER:
       (B) FILING DATE:
       (C) CLASSIFICATION:

 (viii) ATTORNEY/AGENT INFORMATION:
       (A) NAME: Galloway, Norval B

   (ix) TELECOMMUNICATION INFORMATION:
       (A) TELEPHONE: 3128567180
       (B) TELEFAX: 3128564972

(2) INFORMATION FOR SEQ ID NO:1:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 1157 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA (genomic)

-189-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
                                                        AGA  -121

TCTAAAGGCA CAGCGTCTCA TGCTTCGCAC AATGTAAAAC TGCTTCAGAA CCTGGCGAGA  -61

GCTATCCGCG CGGTCTACGG TTAACTGATA CTAAAAGACA ATTCAGCGGG TAACCTTGCA   -1

ATGGTGAGTG GCAGTAAAGC GGGCGTTTCG CCTCATCGCG AAATAGAGGT AATGAGACAA   60

TCCATTGACG ATCACCTGGC TGGCCTGTTA CCTGAAACCG ACAGCCAGGA TATCGTCAGC  120

CTTGCGATGC GTGAAGGCGT CATGGCACCC GGTAAACGGA TCCGTCCGCT GCTGATGCTG  180

CTGGCCGCCC GCGACCTCCG CTACCAGGGC AGTATGCCTA CGCTGCTCGA TCTCGCCTGC  240

GCCGTTGAAC TGACCCATAC CGCGTCGCTG ATGCTCGACG ACATGCCCTG CATGGACACC  300

GCCGAGCTGC CCGCGGTCA GCCCACTACC CACAAAAAAT TTGGTGAGAG CGTGGCGATC  360

CTTGCCTCCG TTGGGCTGCT CTCTAAAGCC TTTGGTCTGA TCGCCGCCAC CGGCGATCTG  420

CCGGGGGAGA GGCGTGCCCA GGCGGTCAAC GAGCTCTCTA CCGCCGTGGG GCTGCAGGGC  480

CTGGTACTGG GGCAGTTTCG CGATCTTAAC GATGCCGCCC TCGACCGTAC CCCTGACGCT  540

ATCCTCAGCA CCAACCACCT CAAGACCGGC ATTCTGTTCA GCGCGATGCT GCAGATCGTC  600

GCCATTGCTT CCGCCTCGTC GCCGAGCACG CGAGAGACGC TGCACGCCTT CGCCCTCGAC  660

TTCGGCCAGG CGTTTCAACT GCTGGACGAT CTGCGTGACG ATCACCCGGA AACCGGTAAA  720

GATCGCAATA AGGACGCGGG AAAATCGACG CTGGTCAACC GGCTGGGCGC AGACGCGGCC  780

CGGCAAAAGC TGCGCGAGCA TATTGATTCC GCCGACAAAC ACCTCACTTT TGCCTGTCCG  840

CAGGGCGGCG CCATCCGACA GTTTATGCAT CTGTGGTTTG GCCATCACCT TGCCGACTGG  900

TCACCGGTCA TGAAAATCGC CTGATACCGC CCTTTTGGGT TCAAGCAGTA CATAACGATG  960

GAACCACATT ACAGGAGTAG TGATGAATGA AGGACGAGCG CCTTGTTCAG CGTAAGAACG 1020

ATCATCTGGA TATC                                                  1034
```

(2) INFORMATION FOR SEQ ID NO:2:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 307 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

-190-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Val Ser Gly Ser Lys Ala Gly Val Ser Pro His Arg Glu Ile Glu
1               5                   10                  15

Val Met Arg Gln Ser Ile Asp Asp His Leu Ala Gly Leu Leu Pro Glu
            20                  25                  30

Thr Asp Ser Gln Asp Ile Val Ser Leu Ala Met Arg Glu Gly Val Met
            35                  40                  45

Ala Pro Gly Lys Arg Ile Arg Pro Leu Leu Met Leu Leu Ala Ala Arg
    50                  55                  60

Asp Leu Arg Tyr Gln Gly Ser Met Pro Thr Leu Leu Asp Leu Ala Cys
65              70                  75                  80

Ala Val Glu Leu Thr His Thr Ala Ser Leu Met Leu Asp Asp Met Pro
                85                  90                  95

Cys Met Asp Asn Ala Glu Leu Arg Arg Gly Gln Pro Thr Thr His Lys
            100                 105                 110

Lys Phe Gly Glu Ser Val Ala Ile Leu Ala Ser Val Gly Leu Leu Ser
    115                 120                 125

Lys Ala Phe Gly Leu Ile Ala Ala Thr Gly Asp Leu Pro Gly Glu Arg
    130                 135                 140

Arg Ala Gln Ala Val Asn Glu Leu Ser Thr Ala Val Gly Leu Gln Gly
145                 150                 155                 160

Leu Val Leu Gly Gln Phe Arg Asp Leu Asn Asp Ala Ala Leu Asp Arg
                165                 170                 175

Thr Pro Asp Ala Ile Leu Ser Thr Asn His Leu Lys Thr Gly Ile Leu
            180                 185                 190

Phe Ser Ala Met Leu Gln Ile Val Ala Ile Ala Ser Ala Ser Ser Pro
            195                 200                 205

Ser Thr Arg Glu Thr Leu His Ala Phe Ala Leu Asp Phe Gly Gln Ala
    210                 215                 220

Phe Gln Leu Leu Asp Asp Leu Arg Asp Asp His Pro Glu Thr Gly Lys
225                 230                 235                 240

Asp Arg Asn Lys Asp Ala Gly Lys Ser Thr Leu Val Asn Arg Leu Gly
                245                 250                 255
```

-191-

```
    Ala Asp Ala Ala Arg Gln Lys Leu Arg Glu His Ile Asp Ser Ala Asp
                260                 265             270

    Lys His Leu Thr Phe Ala Cys Pro Gln Gly Gly Ala Ile Arg Gln Phe
            275             280                 285

    Met His Leu Trp Phe Gly His His Leu Ala Asp Trp Ser Pro Val Met
        290             295             300

    Lys Ile Ala
    305
```

(2) INFORMATION FOR SEQ ID NO:3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1157 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
                                   AGATCTAAAG GCACAGCGTC TCATGCTTCG  -121
CACAATGTAA AACTGCTTCA GAACCTGGCG AGAGCTATCC GCGCGGTCTA CGGTTAACTG   -61
ATACTAAAAG ACAATTCAGC GGGTAACCTT GCAATGGTGA GTGGCAGTAA AGCGGGCGTC    -1
ATGGCCGAAT TCGAAATAGA GGTAATGAGA CAATCCATTG ACGATCACCT GGCTGGCCTG    60
TTACCTGAAA CCGACAGCCA GGATATCGTC AGCCTTGCGA TGCGTGAAGG CGTCATGGCA   120
CCCGGTAAAC GGATCCGTCC GCTGCTGATG CTGCTGGCCG CCCGCGACCT CCGCTACCAG   180
GGCAGTATGC CTACGCTGCT CGATCTCGCC TGCGCCGTTG AACTGACCCA TACCGCGTCG   240
CTGATGCTCG ACGACATGCC CTGCATGGAC ACCGCCGAGC TGCGCCGCGG TCAGCCCACT   300
ACCCACAAAA AATTTGGTGA GAGCGTGGCG ATCCTTGCCT CCGTTGGGCT GCTCTCTAAA   360
GCCTTTGGTC TGATCGCCGC CACCGGCGAT CTGCCGGGGG AGAGGCGTGC CCAGGCGGTC   420
AACGAGCTCT CTACCGCCGT GGGGCTGCAG GGCCTGGTAC TGGGGCAGTT CGCGATCTT    480
AACGATGCCG CCCTCGACCG TACCCCTGAC GCTATCCTCA GCACCAACCA CCTCAAGACC   540
GGCATTCTGT TCAGCGCGAT GCTGCAGATC GTCGCCATTG CTTCCGCCTC GTCGCCGAGC   600
ACGCGAGAGA CGCTGCACGC CTTCGCCCTC GACTTCGGCC AGGCGTTTCA ACTGCTGGAC   660
```

-192-

GATCTGCGTG ACGATCACCC GGAAACCGGT AAAGATCGCA ATAAGGACGC GGGAAAATCG   720

ACGCTGGTCA ACCGGCTGGG CGCAGACGCG GCCCGGCAAA AGCTGCGCGA GCATATTGAT   780

TCCGCCGACA AACACCTCAC TTTTGCCTGT CCGCAGGGCG GCGCCATCCG ACAGTTTATG   840

CATCTGTGGT TTGGCCATCA CCTTGCCGAC TGGTCACCGG TCATGAAAAT CGCCTGATAC   900

CGCCCTTTTG GGTTCAAGCA GTACATAACG ATGGAACCAC ATTACAGGAG TAGTGATGAA   960

TGAAGGACGA GCGCCTTGTT CAGCGTAAGA ACGATCATCT GGATATC   1007

2) INFORMATION FOR SEQ ID NO:4:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 298 amino acids
            (B) TYPE: amino acid
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: protein

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

        Met Ala Glu Phe Glu Ile Glu Val Met Arg Gln Ser Ile Asp Asp His
        1   5   10   15

        Leu Ala Gly Leu Leu Pro Glu Thr Asp Ser Gln Asp Ile Val Ser Leu
                    20   25   30

        Ala Met Arg Glu Gly Val Met Ala Pro Gly Lys Arg Ile Arg Pro Leu
                35   40   45

        Leu Met Leu Leu Ala Ala Arg Asp Leu Arg Tyr Gln Gly Ser Met Pro
            50   55   60

        Thr Leu Leu Asp Leu Ala Cys Ala Val Glu Leu Thr His Thr Ala Ser
        65   70   75   80

        Leu Met Leu Asp Asp Met Pro Cys Met Asp Asn Ala Glu Leu Arg Arg
                        85   90   95

        Gly Gln Pro Thr Thr His Lys Lys Phe Gly Glu Ser Val Ala Ile Leu
                        100   105   110

        Ala Ser Val Gly Leu Leu Ser Lys Ala Phe Gly Leu Ile Ala Ala Thr
                    115   120   125

        Gly Asp Leu Pro Gly Glu Arg Arg Ala Gln Ala Val Asn Glu Leu Ser
                130   135   140

-193-

```
Thr Ala Val Gly Leu Gln Gly Leu Val Leu Gly Gln Phe Arg Asp Leu
145             150                 155                 160

Asn Asp Ala Ala Leu Asp Arg Thr Pro Asp Ala Ile Leu Ser Thr Asn
                165             170                 175

His Leu Lys Thr Gly Ile Leu Phe Ser Ala Met Leu Gln Ile Val Ala
            180             185                 190

Ile Ala Ser Ala Ser Ser Pro Ser Thr Arg Glu Thr Leu His Ala Phe
        195             200                 205

Ala Leu Asp Phe Gly Gln Ala Phe Gln Leu Leu Asp Asp Leu Arg Asp
    210             215                 220

Asp His Pro Glu Thr Gly Lys Asp Arg Asn Lys Asp Ala Gly Lys Ser
225             230                 235                 240

Thr Leu Val Asn Arg Leu Gly Ala Asp Ala Ala Arg Gln Lys Leu Arg
            245             250                 255

Glu His Ile Asp Ser Ala Asp Lys His Leu Thr Phe Ala Cys Pro Gln
            260             265                 270

Gly Gly Ala Ile Arg Gln Phe Met His Leu Trp Phe Gly His His Leu
        275             280                 285

Ala Asp Trp Ser Pro Val Met Lys Ile Ala
    290             295
```

(2) INFORMATION FOR SEQ ID NO:5:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1198 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
                                    GATTG AGGATCTGCA      -1

ATGAGCCAAC CGCCGCTGCT TGACCACGCC ACGCAGACCA TGGCCAACGG CTCGAAAAGT    60

TTTGCCACCG CTGCGAAGCT GTTCGACCCG GCCACCCGCC GTAGCGTGCT GATGCTCTAC   120

ACCTGGTGCC GCCACTGCGA TGACGTCATT GACGACCAGA CCCACGGCTT CGCCAGCGAG   180

GCCGCGGCGG AGGAGGAGGC CACCCAGCGC CTGGCCCGGC TGCGCACGCT GACCCTGGCG   240
```

```
GCGTTTGAAG GGGCCGAGAT GCAGGATCCG GCCTTCGCTG CCTTTCAGGA GGTGGCGCTG   300

ACCCACGGTA TTACGCCCCG CATGGCGCTC GATCACCTCG ACGGCTTTGC GATGGACGTG   360

GCTCAGACCC GGTATGTCAC CTTTGAGGAT ACGCTGCGCT ACTGCTATCA CGTGGCGGGC   420

GTGGTGGGTC TGATGATGGC CAGGGTGATG GGCGTGCGGG ATGAGCGGGT GCTGGATCGC   480

GCCTGCGATC TGGGGCTGGC CTTCCAGCTG ACGAATATGG CCCGGGATAT TATTGACGAT   540

GCGGCTATTG ACCGCTGCTA TCTGCCCGCC GAGTGGCTGC AGGATGCCGG CTGGCCCCG   600

GAGAACTATG CCGCGCGGGA GAATCGCCCC GCGCTGGCGC GGTGGCGGAG GCTTATTGAT   660

GCCGCAGAGC CGTACTACAT CTCCTCCCAG GCCGGGCTAC ACGATCTGCG GCGGCGCTCC   720

GCGTGGGCGA TCGCCACCGC CCGCAGCGTC TACCGGGAGA TCGGTATTAA GGTAAAAGCG   780

GCGGGAGGCA GCGCCTGGGA TCGCCGCCAG CACACCAGCA AAGGTGAAAA AATTGCCATG   840

CTGATGGCGG CACCGGGGCA GGTTATTCGG GCGAAGACGA CGAGGGTGAC GCCGCGTCCG   900

GCCGGTCTTT GGCAGCGTCC CGTTTAGGCG GGCGGCCATG ACGTTCACGC AGGATCGCCT   960

GTAGGTCGGC AGGCTTGCGG GCGTAAATAA AACCGAAGGA GACGCAGCCC TCCCGGCCGC  1020

GCACCGCGTG GTGCAGGCGG TGGGCGACGT AGAGCCGCTT CAGGTAGCCC CGGCGCGGGA  1080

TCCAGTGGAA GGGCCAGCGC TGATGCACCA GACCGTCGTG CACCAGGAAG TAGAGCAGGC  1140

CATAGACCGT CATGCCGCAG CCAATCCACT GCAGGGGCCA AAC                     1183
```

(2) INFORMATION FOR SEQ ID NO:6:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 308 amino acids
       (B) TYPE: amino acid
       (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ser Gln Pro Pro Leu Leu Asp His Ala Thr Gln Thr Met Ala Asn
1               5                   10                  15

Gly Ser Lys Ser Phe Ala Thr Ala Ala Lys Leu Phe Asp Pro Ala Thr
                20                  25                  30

Arg Arg Ser Val Leu Met Leu Tyr Thr Trp Cys Arg His Cys Asp Asp
                35                  40                  45
```

-195-

```
Val Ile Asp Asp Gln Thr His Gly Phe Ala Ser Glu Ala Ala Ala Glu
    50              55              60

Glu Glu Ala Thr Gln Arg Leu Ala Arg Leu Arg Thr Leu Thr Leu Ala
65              70              75              80

Ala Phe Glu Gly Ala Glu Met Gln Asp Pro Ala Phe Ala Ala Phe Gln
            85              90              95

Glu Val Ala Leu Thr His Gly Ile Thr Pro Arg Met Ala Leu Asp His
            100             105             110

Leu Asp Gly Phe Ala Met Asp Val Ala Gln Thr Arg Tyr Val Thr Phe
        115             120             125

Glu Asp Thr Leu Arg Tyr Cys Tyr His Val Ala Gly Val Val Gly Leu
    130             135             140

Met Met Ala Arg Val Met Gly Val Arg Asp Glu Arg Val Leu Asp Arg
145             150             155             160

Ala Cys Asp Leu Gly Leu Ala Phe Gln Leu Thr Asn Met Ala Arg Asp
            165             170             175

Ile Ile Asp Asp Ala Ala Ile Asp Arg Cys Tyr Leu Pro Ala Glu Trp
            180             185             190

Leu Gln Asp Ala Gly Leu Ala Pro Glu Asn Tyr Ala Ala Arg Glu Asn
        195             200             205

Arg Pro Ala Leu Ala Arg Trp Arg Arg Leu Ile Asp Ala Ala Glu Pro
    210             215             220

Tyr Tyr Ile Ser Ser Gln Ala Gly Leu His Asp Leu Arg Arg Arg Ser
225             230             235             240

Ala Trp Ala Ile Ala Thr Ala Arg Ser Val Tyr Arg Glu Ile Gly Ile
            245             250             255

Lys Val Lys Ala Ala Gly Gly Ser Ala Trp Asp Arg Arg Gln His Thr
        260             265             270

Ser Lys Gly Glu Lys Ile Ala Met Leu Met Ala Ala Pro Gly Gln Val
    275             280             285

Ile Arg Ala Lys Thr Thr Arg Val Thr Pro Arg Pro Ala Gly Leu Trp
290             295             300

Gln Arg Pro Val
305
```

-196-

(2) INFORMATION FOR SEQ ID NO:7:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 1518 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
                                                    TAAACC      -1

ATGGAAAAAA CCGTTGTGAT TGGCGCAGGC TTTGGTGGCC TGGCGCTGGC GATTCGCCTG      60

CAGGCGGCAG GGATCCCAAC CGTACTGCTG GAGCAGCGGG ACAAGCCCGG CGGTCGGGCC     120

TACGTCTGGC ATGACCAGGG CTTTACCTTT GACGCCGGGC CGACGGTGAT CACCGATCCT     180

ACCGCGCTTG AGGCGCTGTT CACCCTGGCC GGCAGGCGCA TGGAGGATTA CGTCAGGCTG     240

CTGCCGGTAA AACCCTTCTA CCGACTCTGC TGGGAGTCCG GAAGACCCT CGACTATGCT     300

AACGACAGCT TCGAGCTTGA GGCGCAGATT ACCCAGTTCA ACCCCGCGA CGTCGAGGGC     360

TACCGGCGCT TTCTGGCTTA CTCCCAGGCG GTATTCCAGG AGGGATATTT GCGCCTCGGC     420

AGCGTGCCGT TCCTCTCTTT TCGCGACATG CTGCGCGCCG GCCGCAGCT GCTTAAGCTC     480

CAGGCGTGGC AGAGCGTCTA CCAGTCGGTT TCGCGCTTTA TTGAGGATGA GCATCTGCGG     540

CAGGCCTTCT CGTTCCACTC CCTGCTGGTA GGCGGCAACC CCTTCACCAC CTCG0TCCATC     600

TACACCCTGA TCCACGCCCT TGAGCGGGAG TGGGGGGTCT GGTTCCCTGA GGGCGGCACC     660

GGGGCGCTGG TGAACGGCAT GGTGAAGCTG TTTACCGATC TGGGCGGGGA GATCGAACTC     720

AACGCCCGGG TCGAAGAGCT GGTGGTGGCC GATAACCGCG TAAGCCAGGT CCGGCTCGCG     780

GATGGTCGGA TCTTTGACAC CGACGCCGTA GCCTCGAACG CTGACGTGGT GAACACCTAT     840

AAAAAGCTGC TCGGCACCAT ACCGGTGGGG CAGAAGCGGG CGGCACGGCT GGAGCGCAAG     900

AGCATGAGCA ACTCGCTGTT TGTGCTCTAC TTCGGCCTGA ACCAGCCTCA TTCCCAGCTG     960

GCGCACCATA CCATCTGTTT TGGTCCCCGC TACCGGGAGC TGATCGACGA GATCTTTACC    1020

GGCAGCGCGC TGGCGGATGA CTTCTCGCTC TACCTGCACT CGCCCTGCGT GACCGATCCC    1080

TCGCTCGCGC CTCCCCCGTG CGCCAGCTTC TACGTGCTGG CCCCGGTGCC GCATCTTGGC    1140

AACGCGCCGC TGGACTGGGC GCAGGAGGGG CCGAAGCTGC GCGACCGCAT CTTTGACTAC    1200
```

-197-

CTTGAAGAGC GCTATATGCC CGGCCTGCGT AGCCAGCTGG TGACCCAGCG GATCTTTACC   1260

CGGCAGACTT CACGACACGC TTGGATCGCG ATCTTGGGAT CGCTTTTCAT CGAGCCGCCT   1320

TCGTTGACCC AAGGCTTGTT CGCCGCAAAC GCGACACGAC ATTCAAACCT CTACCTGGTG   1380

GCCGCAGGTA CTCACCCTGG CGCGGGCATT CCTGGCGTAG TGGGCCTCGC CGAAAGCACC   1440

GCCAGCCTGA TGATTGAGGA TCTGCAATGA GCCAACCGCC GCTGCTTGAC CACGCCACGC   1500

AGACCATGGC CA   1512

(2) INFORMATION FOR SEQ ID NO:8:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 489 amino acids
       (B) TYPE: amino acid
       (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Lys Lys Thr Val Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
1   5   10   15

Ala Ile Arg Leu Gln Ala Ala Gly Ile Pro Thr Val Leu Leu Glu Gln
            20   25   30

Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Trp His Asp Gln Gly Phe
        35   40   45

Thr Phe Asp Ala Gly Pro Thr Val Ile Thr Asp Pro Thr Ala Leu Glu
    50   55   60

Ala Leu Phe Thr Leu Ala Gly Arg Arg Met Glu Asp Tyr Val Arg Leu
65   70   75   80

Leu Pro Val Lys Pro Phe Tyr Arg Leu Cys Trp Glu Ser Gly Lys Thr
            85   90   95

Leu Asp Tyr Ala Asn Asp Ser Phe Glu Leu Glu Ala Gln Ile Thr Gln
            100   105   110

Phe Asn Pro Arg Asp Val Glu Gly Tyr Arg Arg Phe Leu Ala Tyr Ser
        115   120   125

Gln Ala Val Phe Gln Glu Gly Tyr Leu Arg Leu Gly Ser Val Pro Phe
    130   135   140

-198-

Leu Ser Phe Arg Asp Met Leu Arg Ala Gly Pro Gln Leu Leu Lys Leu
145   150   155   160

Gln Ala Trp Gln Ser Val Tyr Gln Ser Val Ser Arg Phe Ile Glu Asp
            165   170   175

Glu His Leu Arg Gln Ala Phe Ser Phe His Ser Leu Leu Val Gly Gly
            180   185   190

Asn Pro Phe Thr Thr Ser Ser Ile Tyr Thr Leu Ile His Ala Leu Glu
            195   200   205

Arg Glu Trp Gly Val Trp Phe Pro Glu Gly Gly Thr Gly Ala Leu Val
210   215   220

Asn Gly Met Val Lys Leu Phe Thr Asp Leu Gly Gly Glu Ile Glu Leu
225   230   235   240

Asn Ala Arg Val Glu Glu Leu Val Val Ala Asp Asn Arg Val Ser Gln
            245   250   255

Val Arg Leu Ala Asp Gly Arg Ile Phe Asp Thr Asp Ala Val Ala Ser
            260   265   270

Asn Ala Asp Val Val Asn Thr Tyr Lys Lys Leu Leu Gly Thr Ile Pro
            275   280   285

Val Gly Gln Lys Arg Ala Ala Arg Leu Glu Arg Lys Ser Met Ser Asn
290   295   300

Ser Leu Phe Val Leu Tyr Phe Gly Leu Asn Gln Pro His Ser Gln Leu
305   310   315   320

Ala His His Thr Ile Cys Phe Gly Pro Arg Tyr Arg Glu Leu Ile Asp
            325   330   335

Glu Ile Phe Thr Gly Ser Ala Leu Ala Asp Asp Phe Ser Leu Tyr Leu
            340   345   350

His Ser Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Pro Pro Cys Ala
            355   360   365

Ser Phe Tyr Val Leu Ala Pro Val Pro His Leu Gly Asn Ala Pro Leu
370   375   380

Asp Trp Ala Gln Glu Gly Pro Lys Leu Arg Asp Arg Ile Phe Asp Tyr
385   390   395   400

Leu Glu Glu Arg Tyr Met Pro Gly Leu Arg Ser Gln Leu Val Thr Gln
            405   410   415

Arg Ile Phe Thr Arg Gln Thr Ser Arg His Ala Trp Ile Ala Ile Leu
            420   425   430

-199-

```
     Gly Ser Leu Phe Ile Glu Pro Pro Ser Leu Thr Gln Gly Leu Phe Ala
             435             440             445

     Ala Asn Ala Thr Arg His Ser Asn Leu Tyr Leu Val Ala Ala Gly Thr
             450             455             460

     His Pro Gly Ala Gly Ile Pro Gly Val Val Gly Leu Ala Glu Ser Thr
     465             470             475             480

     Ala Ser Leu Met Ile Glu Asp Leu Gln
                     485
```

(2) INFORMATION FOR SEQ ID NO:9:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 1522 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
                                                    GAGGTCGACG    -1
ATGGAAAAAA CCGTTGTGAT TGGCGCAGGC TTTGGTGGCC TGGCGCTGGC GATTCGCCTG    60

CAGGCGGCAG GGATCCCAAC CGTACTGCTG GAGCAGCGGG ACAAGCCCGG CGGTCGGGCC   120

TACGTCTGGC ATGACCAGGG CTTTACCTTT GACGCCGGGC CGACGGTGAT CACCGATCCT   180

ACCGCGCTTG AGGCGCTGTT CACCCTGGCC GGCAGGCGCA TGGAGGATTA CGTCAGGCTG   240

CTGCCGGTAA AACCCTTCTA CCGACTCTGC TGGGAGTCCG GAAGACCCT CGACTATGCT   300

AACGACAGCT TCGAGCTTGA GGCGCAGATT ACCCAGTTCA ACCCCCGCGA CGTCGAGGGC   360

TACCGGCGCT TTCTGGCTTA CTCCCAGGCG GTATTCCAGG AGGGATATTT GCGCCTCGGC   420

AGCGTGCCGT TCCTCTCTTT TCGCGACATG CTGCGCGCCG GCCGCAGCT GCTTAAGCTC   480

CAGGCGTGGC AGAGCGTCTA CCAGTCGGTT TCGCGCTTTA TTGAGGATGA GCATCTGCGG   540

CAGGCCTTCT CGTTCCACTC CCTGCTGGTA GGCGGCAACC CCTTCACCAC CTCGTCCATC   600

TACACCCTGA TCCACGCCCT TGAGCGGGAG TGGGGGGTCT GGTTCCCTGA GGGCGGCACC   660

GGGGCGCTGG TGAACGGCAT GGTGAAGCTG TTTACCGATC TGGGCGGGGA GATCGAACTC   720

AACGCCCGGG TCGAAGAGCT GGTGGTGGCC GATAACCGCG TAAGCCAGGT CCGGCTCGCG   780
```

-200-

GATGGTCGGA TCTTTGACAC CGACGCCGTA GCCTCGAACG CTGACGTGGT GAACACCTAT 840

AAAAAGCTGC TCGGCACCAT ACCGGTGGGG CAGAAGCGGG CGGCACGGCT GGAGCGCAAG 900

AGCATGAGCA ACTCGCTGTT TGTGCTCTAC TTCGGCCTGA ACCAGCCTCA TTCCCAGCTG 960

GCGCACCATA CCATCTGTTT TGGTCCCCGC TACCGGGAGC TGATCGACGA GATCTTTACC 1020

GGCAGCGCGC TGGCGGATGA CTTCTCGCTC TACCTGCACT CGCCCTGCGT GACCGATCCC 1080

TCGCTCGCGC CTCCCCCGTG CGCCAGCTTC TACGTGCTGG CCCCGGTGCC GCATCTTGGC 1140

AACGCGCCGC TGGACTGGGC GCAGGAGGGG CCGAAGCTGC GCGACCGCAT CTTTGACTAC 1200

CTTGAAGAGC GCTATATGCC CGGCCTGCGT AGCCAGCTGG TGACCCAGCG GATCTTTACC 1260

CGGCAGACTT CACGACACGC TTGGATCGCG ATCTTGGGAT CGCTTTTCAT CGAGCCGCCT 1320

TCGTTGACCC AAGGCTTGTT CGCCGCAAAC GCGACACGAC ATTCAAACCT CTACCTGGTG 1380

GCCGCAGGTA CTCACCCTGG CGCGGGCATT CCTGGCGTAG TGGGCCTCGC CGAAAGCACC 1440

GCCAGCCTGA TGATTGAGGA TCTGCAATGA GCCAACCGCC GCTGCTTGAC CACGCCACGT 1500

CGACCATGGC CA   1512

(2) INFORMATION FOR SEQ ID NO:10:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 177 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGGCTTCCT CAGTTCTTTC CTCTGCAGCA GTTGCCACCC GCAGCAATGT TGCTCAAGCT   60

AACATGGTGG CGCCTTTCAC TGGCCTTAAG TCAGCTGCCT CATTCCCTGT TTCAAGGAAG 120

CAAAACCTTG ACATCACTTC CATTGCCAGC AACGGCGGAA GAGTGCAATG CATGCAG   177

-201-

(2) INFORMATION FOR SEQ ID NO:11:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1235 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
                                          CAGGGAGTG AGAGCGTATC    -1

GTGAGGGATC TGATTTTAGT CGGCCGCGGC CTGGCCAACG GGCTGATCGC CTGGCGTCTG    60

CGCCAGCGCT ACCCGCAGCT TAACCTGCTG CTGATCGAGG CCGGGGAGCA GCCCGGCGGG   120

AACCATACCT GGTCATTCCA TGAAGACGAT CTGACTCCCG GCAGCACGC CTGGCTGGCC   180

CCGCTGGTGG CCCACGCCTG GCCGGGCTAT GAGGTGCAGT TTCCCGATCT TCGCCGTCGC   240

CTCGCGCGCG GCTACTACTC CATTACCTCA GAGCGCTTTG CCGAGGCCCT GCATCAGGCG   300

CTGGGGGAGA ACATCTGGCT AAACTGTTCG GTGAGCGAGG TGTTACCCAA TAGCGTGCGC   360

CTTGCCAACG GTGAGGCGCT GCTTGCCGGA GCGGTGATTG ACGGACGCGG CGTGACCGCC   420

AGTTCGGCGA TGCAAACCGG CTATCAGCTC TTTCTTGGTC AGCAGTGGCG GCTGACACAG   480

CCCCACGGCC TGACCGTACC GATCCTGATG GATGCCACGG TGGCGCAGCA GCAGGGCTAT   540

CGCTTTGTCT ACACGCTGCC GCTCTCCGCC GACACGCTGC TGATCGAGGA TACGCGCTAC   600

GCCAATGTCC CGCAGCGTGA TGATAATGCC CTACGCCAGA CGGTTACCGA CTATGCTCAC   660

AGCAAAGGGT GGCAGCTGGC CCAGCTTGAA CGCGAGGAGA CCGGCTGTCT GCCGATTACC   720

TGGCGGGTGA CATCCAGGCT CTGTGGGCCG ATGCGCCGGC GTGCCGCGTC GGGAATGCGG   780

GCTGGGCTAT TTCACCCTAC CACTGGCTAT TCGCTGCCGC TGGCGGTGGC CCTTGCCGAC   840

GCGATTGCCG ACAGCCCGCG GCTGGGCAGC GTTCCGCTCT ATCAGCTCAC CCGGCAGTTT   900

GCCGAACGCC ACTGGCGCAG GCAGGGATTC TTCCGCCTGC TGAACCGGAT GCTTTTCCTG   960

GCCGGGCGCG AGGAGAACCG CTGGCGGGTG ATGCAGCGCT TTTATGGGCT GCCGGAGCCC  1020

ACCGTAGAGC GCTTTTACGC CGGTCGGCTC TCTCTCTTTG ATAAGGCCCG CATTTTGACG  1080

GGCAAGCCAC CGGTTCCGCT GGCGAAGTCT GGCGGGCGGC GCTGAACCAT TTTCCTGACA  1140
```

-202-

GACGAGATAA AGGATGAAAA AAACCGTTGT GATTGGCGCA GGCTTTGGTG GCCTGGCGCT 1200

GGCGATTCGC CTGCAG   1216

(2) INFORMATION FOR SEQ ID NO:12:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1235 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
                                        CAGGGAGTG AGAGGCGCGC    -1

ATGCGGGATC TGATTTTAGT CGGCCGCGGC CTGGCCAACG GGCTGATCGC CTGGCGTCTG    60

CGCCAGCGCT ACCCGCAGCT TAACCTGCTG CTGATCGAGG CCGGGGAGCA GCCCGGCGGG   120

AACCATACCT GGTCATTCCA TGAAGACGAT CTGACTCCCG GCAGCACGC CTGGCTGGCC   180

CCGCTGGTGG CCCACGCCTG GCCGGGCTAT GAGGTGCAGT TTCCCGATCT TCGCCGTCGC   240

CTCGCGCGCG GCTACTACTC CATTACCTCA GAGCGCTTTG CCGAGGCCCT GCATCAGGCG   300

CTGGGGGAGA ACATCTGGCT AAACTGTTCG GTGAGCGAGG TGTTACCCAA TAGCGTGCGC   360

CTTGCCAACG GTGAGGCGCT GCTTGCCGGA GCGGTGATTG ACGGACGCGG CGTGACCGCC   420

AGTTCGGCGA TGCAAACCGG CTATCAGCTC TTTCTTGGTC AGCAGTGGCG GCTGACACAG   480

CCCCACGGCC TGACCGTACC GATCCTGATG GATGCCACGG TGGCGCAGCA GCAGGGCTAT   540

CGCTTTGTCT ACACGCTGCC GCTCTCCGCC GACACGCTGC TGATCGAGGA TACGCGCTAC   600

GCCAATGTCC CGCAGCGTGA TGATAATGCC CTACGCCAGA CGGTTACCGA CTATGCTCAC   660

AGCAAAGGGT GGCAGCTGGC CCAGCTTGAA CGCGAGGAGA CCGGCTGTCT GCCGATTACC   720

TGGCGGGTGA CATCCAGGCT CTGTGGGCCG ATGCGCCGGC GTGCCGCGTC GGGAATGCGG   780

GCTGGGCTAT TTCACCCTAC CACTGGCTAT TCGCTGCCGC TGGCGGTGGC CCTTGCCGAC   840

GCGATTGCCG ACAGCCCGCG GCTGGGCAGC GTTCCGCTCT ATCAGCTCAC CCGGCAGTTT   900

GCCGAACGCC ACTGGCGCAG GCAGGGATTC TTCCGCCTGC TGAACCGGAT GCTTTTCCTG   960

GCCGGGCGCG AGGAGAACCG CTGGCGGGTG ATGCAGCGCT TTTATGGGCT GCCGGAGCCC  1020
```

-203-

ACCGTAGAGC GCTTTTACGC CGGTCGGCTC TCTCTCTTTG ATAAGGCCCG CATTTTGACG 1080

GGCAAGCCAC CGGTTCCGCT GGCGAAGTCT GGCGGGCGGC GCTGAACCAT TTTCCTGACA 1140

GACGAGATAA AGGGATCCGA TGACCGTTGT GATTGGCGCA GGCTTTGGTG GCCTGGCGCT 1200

GGCGATTCGC CTGCAG   1216

(2) INFORMATION FOR SEQ ID NO:13:

     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 374 amino acids
         (B) TYPE: amino acid
         (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

    Met Arg Asp Leu Ile Leu Val Gly Arg Gly Leu Ala Asn Gly Leu Ile
    1   5   10   15

    Ala Trp Arg Leu Arg Gln Arg Tyr Pro Gln Leu Asn Leu Leu Leu Ile
                20   25   30

    Glu Ala Gly Glu Gln Pro Gly Gly Asn His Thr Trp Ser Phe His Glu
                35   40   45

    Asp Asp Leu Thr Pro Gly Gln His Ala Trp Leu Ala Pro Leu Val Ala
        50   55   60

    His Ala Trp Pro Gly Tyr Glu Val Gln Phe Pro Asp Leu Arg Arg Arg
    65   70   75   80

    Leu Ala Arg Gly Tyr Tyr Ser Ile Thr Ser Glu Arg Phe Ala Glu Ala
                    85   90   95

    Leu His Gln Ala Leu Gly Glu Asn Ile Trp Leu Asn Cys Ser Val Ser
                100   105   110

    Glu Val Leu Pro Asn Ser Val Arg Leu Ala Asn Gly Glu Ala Leu Leu
                115   120   125

    Ala Gly Ala Val Ile Asp Gly Arg Gly Val Thr Ala Ser Ser Ala Met
        130   135   140

    Gln Thr Gly Tyr Gln Leu Phe Leu Gly Gln Gln Trp Arg Leu Thr Gln
    145   150   155   160

-204-

```
Pro His Gly Leu Thr Val Pro Ile Leu Met Asp Ala Thr Val Ala Gln
             165         170             175

Gln Gln Gly Tyr Arg Phe Val Tyr Thr Leu Pro Leu Ser Ala Asp Thr
             180         185             190

Leu Leu Ile Glu Asp Thr Arg Tyr Ala Asn Val Pro Gln Arg Asp Asp
             195         200             205

Asn Ala Leu Arg Gln Thr Val Thr Asp Tyr Ala His Ser Lys Gly Trp
    210             215         220

Gln Leu Ala Gln Leu Glu Arg Glu Glu Thr Gly Cys Leu Pro Ile Thr
225             230         235             240

Trp Arg Val Thr Ser Arg Leu Cys Gly Pro Met Arg Arg Arg Ala Ala
                245         250             255

Ser Gly Met Arg Ala Gly Leu Phe His Pro Thr Thr Gly Tyr Ser Leu
            260         265             270

Pro Leu Ala Val Ala Leu Ala Asp Ala Ile Ala Asp Ser Pro Arg Leu
        275         280             285

Gly Ser Val Pro Leu Tyr Gln Leu Thr Arg Gln Phe Ala Glu Arg His
    290             295         300

Trp Arg Arg Gln Gly Phe Phe Arg Leu Leu Asn Arg Met Leu Phe Leu
305             310             315             320

Ala Gly Arg Glu Glu Asn Arg Trp Arg Val Met Gln Arg Phe Tyr Gly
            325             330             335

Leu Pro Glu Pro Thr Val Glu Arg Phe Tyr Ala Gly Arg Leu Ser Leu
            340             345             350

Phe Asp Lys Ala Arg Ile Leu Thr Gly Lys Pro Pro Val Pro Leu Ala
        355             360             365

Lys Ser Gly Gly Arg Arg
370
```

(2) INFORMATION FOR SEQ ID NO:14:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 947 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

-205-

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
                                        TGAG TCGCAGGAGC GCGCACCGCT      -1
ATGCTAGTAA ATAGTTTAAT CGTCATCTTG ACCGTTATTG CGATGGAGGG CATCGCCGCG      60
TTTACCCACC GCTACATTAT GCACGGCTGG GGATGGCGCT GGCATGAGCC ACACCATACC     120
CCGCGCAAGG GCGTATTTGA GCTAAACGAT CTCTTTGCGG TGGTGTTTGC CGGGGTGGCT     180
ATCGCGCTGA TTGCCGTGGG CACGGCGGGC GTTTGGCCCC TGCAGTGGAT TGGCTGCGGC     240
ATGACGGTCT ATGGCCTGCT CTACTTCCTG GTGCACGACG GTCTGGTGCA TCAGCGCTGG     300
CCCTTCCACT GGATCCCGCG CCGGGGCTAC CTGAAGCGGC TCTACGTCGC CCACCGCCTG     360
CACCACGCGG TGCGCGGCCG GGAGGGCTGC GTCTCCTTCG GTTTTATTTA CGCCCGCAAG     420
CCTGCCGACC TACAGGCGAT CCTGCGTGAA CGTCATGGCC GCCCGCCTAA ACGGGACGCT     480
GCCAAAGACC GGCCGGACGC GGCGTCACCC TCGTCGTCTT CGCCCGAATA ACCTGCCCCG     540
GTGCCGCCAT CAGCATGGCA ATTTTTTCAC CTTTGCTGGT GTGCTGGCGG CGATCCCAGG     600
CGCTGCCTCC CGCCGCTTTT ACCTTAATAC CGATCTCCCG GTAGACGCTG CGGGCGGTGG     660
CGATCGCCCA CGCGGAGCGC CGCCGCAGAT CGTGTAGCCC GGCCTGGGAG GAGATGTAGT     720
ACGGCTCTGC GGCATCAATA AGCCTCCGCC ACCGCGCCAG CGCGGGGCGA TTCTCCCGCG     780
CGGCATAGTT CTCCGGGGCC AGCCCGGCAT CCTGCAGCCA CTCGGCGGGC AGATAGCAGC     840
GGTCAATAGC CGCATCGTCA ATAATATCCC GGGCCATATT CGTCAGCTGG AAGGCCAGCC     900
CCAGATCGCA GGCGCGATCC AGC                                             923
```

(2) INFORMATION FOR SEQ ID NO:15:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 947 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
                                    TGAG TCGCAGGAGC GCGCACCGCC    -1

ATGGTACTAA ATAGTTTAAT CGTCATCTTG ACCGTTATTG CGATGGAGGG CATCGCCGCG    60

TTTACCCACC GCTACATTAT GCACGGCTGG GGATGGCGCT GGCATGAGCC ACACCATACC   120

CCGCGCAAGG GCGTATTTGA GCTAAACGAT CTCTTTGCGG TGGTGTTTGC CGGGGTGGCT   180

ATCGCGCTGA TTGCCGTGGG CACGGCGGGC GTTTGGCCCC TGCAGTGGAT TGGCTGCGGC   240

ATGACGGTCT ATGGCCTGCT CTACTTCCTG GTGCACGACG GTCTGGTGCA TCAGCGCTGG   300

CCCTTCCACT GGATCCCGCG CCGGGGCTAC CTGAAGCGGC TCTACGTCGC CCACCGCCTG   360

CACCACGCGG TGCGCGGCCG GGAGGGCTGC GTCTCCTTCG GTTTTATTTA CGCCCGCAAG   420

CCTGCCGACC TACAGGCGAT CCTGCGTGAA CGTCATGGCC GCCCGCCTAA ACGGGACGCT   480

GCCAAAGACC GGCCGGACGC GGCGTCACCC TCGTCGTCTT CGCCCGAATA ACCTGCCCCG   540

GTGCCGCCAT CAGCATGGCA ATTTTTTCAC CTTTGCTGGT GTGCTGGCGG CGATCCCAGG   600

CGCTGCCTCC CGCCGCTTTT ACCTTAATAC CGATCTCCCG GTAGACGCTG CGGGCGGTGG   660

CGATCGCCCA CGCGGAGCGC CGCCGCAGAT CGTGTAGCCC GGCCTGGGAG GAGATGTAGT   720

ACGGCTCTGC GGCATCAATA AGCCTCCGCC ACCGCGCCAG CGCGGGGCGA TTCTCCCGCG   780

CGGCATAGTT CTCCGGGGCC AGCCCGGCAT CCTGCAGCCA CTCGGCGGGC AGATAGCAGC   840

GGTCAATAGC CGCATCGTCA ATAATATCCC GGGCCATATT CGTCAGCTGG AAGGCCAGCC   900

CCAGATCGCA GGCGCGATCC AGC                                          923
```

(2) INFORMATION FOR SEQ ID NO:16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 176 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Leu Val Asn Ser Leu Ile Val Ile Leu Thr Val Ile Ala Met Glu
1               5                   10                  15
```

-207-

```
        Gly Ile Ala Ala Phe Thr His Arg Tyr Ile Met His Gly Trp Gly Trp
                    20                  25                  30

        Arg Trp His Glu Pro His His Thr Pro Arg Lys Gly Val Phe Glu Leu
                    35                  40                  45

        Asn Asp Leu Phe Ala Val Val Phe Ala Gly Val Ala Ile Ala Leu Ile
            50                  55                  60

        Ala Val Gly Thr Ala Gly Val Trp Pro Leu Gln Trp Ile Gly Cys Gly
        65                  70                  75                  80

        Met Thr Val Tyr Gly Leu Leu Tyr Phe Leu Val His Asp Gly Leu Val
                        85                  90                  95

        His Gln Arg Trp Pro Phe His Trp Ile Pro Arg Arg Gly Tyr Leu Lys
                    100                 105                 110

        Arg Leu Tyr Val Ala His Arg Leu His His Ala Val Arg Gly Arg Glu
                115                 120                 125

        Gly Cys Val Ser Phe Gly Phe Ile Tyr Ala Arg Lys Pro Ala Asp Leu
            130                 135                 140

        Gln Ala Ile Leu Arg Glu Arg His Gly Arg Pro Pro Lys Arg Asp Ala
        145                 150                 155                 160

        Ala Lys Asp Arg Pro Asp Ala Ala Ser Pro Ser Ser Ser Ser Pro Glu
                        165                 170                 175
```

(2) INFORMATION FOR SEQ ID NO:17:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 399 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
        Met Ser His Phe Ala Ile Val Ala Pro Pro Leu Tyr Ser His Ala Val
        1               5                   10                  15

        Ala Leu His Ala Leu Ala Leu Glu Met Ala Gln Arg Gly His Arg Val
                    20                  25                  30

        Thr Phe Leu Thr Gly Asn Val Ala Ser Leu Ala Glu Gln Glu Thr Glu
                    35                  40                  45
```

-208-

```
Arg Val Ala Phe Tyr Pro Leu Pro Ala Ser Val Gln Gln Ala Gln Arg
    50              55              60

Asn Val Gln Gln Gln Ser Asn Gly Asn Leu Leu Arg Leu Ile Ala Ala
65              70              75              80

Met Ser Ser Leu Thr Asp Val Leu Cys Gln Gln Leu Pro Ala Ile Leu
            85              90              95

Gln Arg Leu Ala Val Asp Ala Leu Ile Val Asp Glu Met Glu Pro Ala
        100             105             110

Gly Ser Leu Val Ala Glu Ala Leu Gly Leu Pro Phe Ile Ser Ile Ala
        115             120             125

Cys Ala Leu Pro Val Asn Arg Glu Leu Pro Leu Pro Val Met Pro Phe
    130             135             140

His Tyr Ala Glu Asp Lys Arg Ala Arg Ala Arg Phe Gln Val Ser Glu
145             150             155             160

Arg Ile Tyr Asp Ala Leu Met Tyr Pro His Gly Gln Thr Ile Leu Arg
            165             170             175

His Ala Gln Arg Phe Gly Leu Pro Glu Arg Arg Arg Leu Asp Glu Cys
        180             185             190

Leu Ser Pro Leu Ala Gln Ile Ser Gln Ser Val Pro Ala Leu Asp Phe
    195             200             205

Pro Arg Arg Ala Leu Pro Asn Cys Phe Thr Tyr Val Gly Ala Leu Arg
210             215             220

Tyr Gln Pro Pro Pro Gln Val Glu Arg Ser Pro Arg Ser Thr Pro Arg
225             230             235             240

Ile Phe Ala Ser Leu Gly Thr Leu Gln Gly His Arg Leu Arg Leu Phe
            245             250             255

Gln Lys Ile Ala Arg Ala Cys Ala Ser Val Gly Ala Glu Val Thr Ile
        260             265             270

Ala His Cys Asp Gly Leu Thr Pro Ala Gln Ala Asp Ser Leu Tyr Cys
        275             280             285

Gly Ala Thr Glu Val Val Ser Phe Val Asp Gln Pro Arg Tyr Val Ala
290             295             300

Glu Ala Asn Leu Val Ile Thr His Gly Gly Leu Asn Thr Val Leu Asp
305             310             315             320

Ala Leu Ala Ala Ala Thr Pro Val Leu Ala Val Pro Leu Ser Phe Asp
            325             330             335
```

```
Gln Pro Ala Val Ala Ala Arg Leu Val Tyr Asn Gly Leu Gly Arg Arg
            340             345             350
```

```
Val Ser Arg Phe Ala Arg Gln Gln Thr Leu Ala Asp Glu Ile Ala Gln
        355             360             365
```

```
Leu Leu Gly Asp Glu Thr Leu His Gln Arg Val Ala Thr Ala Arg Gln
        370             375             380
```

```
Gln Leu Asn Asp Ala Gly Gly Thr Pro Arg Cys Gly Asp Pro Asp
385             390             395
```

(2) INFORMATION FOR SEQ ID NO:18:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 39 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCAGCGGGTA ACCTTGCCAT GGGGAGTGGC AGTAAAGCG   39

(2) INFORMATION FOR SEQ ID NO:19:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTGCAATGGT GA   12

(2) INFORMATION FOR SEQ ID NO:20:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

-210-

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTGCCATGGG GA   12

    (2) INFORMATION FOR SEQ ID NO:21:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 39 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CATGGCGAAA TAGAAGCCAT GGGACAATCC ATTGACGAT   39

    (2) INFORMATION FOR SEQ ID NO:22:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 12 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

AAGTAATGAG AC   12

    (2) INFORMATION FOR SEQ ID NO:23:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 12 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AAGCCATGGG AC   12

-211-

(2) INFORMATION FOR SEQ ID NO:24:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 15 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Met Val Ser Gly Ser Lys Ala Gly Val Ser Pro His Arg Glu Ile
1               5                   10                  15
```


(2) INFORMATION FOR SEQ ID NO:25:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear
    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
Met Ala Glu Phe Glu Ile
1               5
```


(2) INFORMATION FOR SEQ ID NO:26:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 46 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
GATCTAAAAT GAGCCAACCG CCGCTGCTTG ACCACGCCAC GCAGAC   46
```


(2) INFORMATION FOR SEQ ID NO:27:

-212-

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 46 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CATGGTCTGC GTGGCGTGGT CAAGCAGCGG CGGTTGGCTC ATTTTA   46

    (2) INFORMATION FOR SEQ ID NO:28:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 28 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ACAACAAAAT ATAAAAACAA TGTCTTTA   28

    (2) INFORMATION FOR SEQ ID NO:29:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 28 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ACAACAAGAT CTAAAAACAA TGTCTTTA   28

    (2) INFORMATION FOR SEQ ID NO:30:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 17 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

-213-

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

AATTCCCGGG CCATGGC   17

(2) INFORMATION FOR SEQ ID NO:31:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 17 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AATTGCCATG GCCCGGG   17

(2) INFORMATION FOR SEQ ID NO:32:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 29 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AAACCATGGA AAAAACCGTT GTGATTGGC   29

(2) INFORMATION FOR SEQ ID NO:33:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 21 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGCCATGGTC TGCGTGGCGT G   21

-214-

(2) INFORMATION FOR SEQ ID NO:34:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TAAAGGATGA AAAAAACCGT TGTGATTGGC   30

(2) INFORMATION FOR SEQ ID NO:35:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 8 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Lys Lys Thr Val Val Ile Gly
1   5

(2) INFORMATION FOR SEQ ID NO:36:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TAAACCATGG AAAAAACCGT TGTGATTGGC   30

(2) INFORMATION FOR SEQ ID NO:37:

-215-

```
     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 8 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

    Met Glu Lys Thr Val Val Ile Gly
    1               5
```

(2) INFORMATION FOR SEQ ID NO:38:

```
     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 30 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:
```

GAGATAAAGG ATGAAAAAAA CCGTTGTGAT   30

(2) INFORMATION FOR SEQ ID NO:39:

```
     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 30 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:
```

GAGGTCGACG ATGAAAAAAA CCGTTGTGAT   30

(2) INFORMATION FOR SEQ ID NO:40:

```
     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 29 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear
```

-216-

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATGGTCGACG TGGCGTGGTC AAGCAGCGG   29

(2) INFORMATION FOR SEQ ID NO:41:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GAGATAAAGG ATGAAAAAAA CCGTTGTGAT   30

(2) INFORMATION FOR SEQ ID NO:42:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

    Met Lys Lys Thr Val Val
    1   5

(2) INFORMATION FOR SEQ ID NO:43:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 22 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CCATGGAAAA AACCGTTGTG AT   22

-217-

(2) INFORMATION FOR SEQ ID NO:44:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

    Met Glu Lys Thr Val Val
    1   5

(2) INFORMATION FOR SEQ ID NO:45:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAGGTCGACG ATGAAAAAAA CCGTTGTGAT   30

(2) INFORMATION FOR SEQ ID NO:46:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

    Met Lys Lys Thr Val Val
    1   5

(2) INFORMATION FOR SEQ ID NO:47:

-218-

```
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 31 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:
```

CCGCTGCTTG ACCACGCCAC GCAGACCATG G   31

(2) INFORMATION FOR SEQ ID NO:48:

```
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 31 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:
```

CCGCTGCTTG ACCACGCCAC GTCGACCATG G   31

(2) INFORMATION FOR SEQ ID NO:49:

```
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 30 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:
```

GACGAGATAA AGCATGCAAA AAACCGTTGT   30

(2) INFORMATION FOR SEQ ID NO:50:

```
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 5 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear
```

-219-

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

    Met Gln Lys Thr Val
    1   5

(2) INFORMATION FOR SEQ ID NO:51:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GACGAGATAA AGGATGAAAA AAACCGTTGT   30

(2) INFORMATION FOR SEQ ID NO:52:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 5 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

    Met Lys Lys Thr Val
    1   5

(2) INFORMATION FOR SEQ ID NO:53:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

-220-

GACGAGATAA AGCATGCAAA AAACCGTTGT   30

(2) INFORMATION FOR SEQ ID NO:54:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 5 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: peptide

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

       Met Gln Lys Thr Val
       1   5

(2) INFORMATION FOR SEQ ID NO:55:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 37 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA (genomic)

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CATGGCTTCC TCAGTTCTTT CCTCTGCAGC AGTTGCC   37

(2) INFORMATION FOR SEQ ID NO:56:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 37 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA (genomic)

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGGTGGCAAC TGCTGCAGAG GAAAGAACTG AGGAAGC   37

-221-

(2) INFORMATION FOR SEQ ID NO:57:

  (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 34 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: DNA (genomic)

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ACCCGCAGCA ATGTTGCTCA AGCTAACATG GTGG 34

(2) INFORMATION FOR SEQ ID NO:58:

  (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 32 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: DNA (genomic)

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CGCCACCATG TTAGCTTGAG CAACATTGCT GC 32

(2) INFORMATION FOR SEQ ID NO:59:

  (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 50 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: DNA (genomic)

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CGCCTTTCAC TGGCCTTAAG TCAGCTGCCT CATTCCCTGT TTCAAGGAAG 50

(2) INFORMATION FOR SEQ ID NO:60:

-222-

```
     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 64 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: single
         (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:
```

TTTGCTTCCT TGAAACAGGG AATGAGGCAG CGAATGAGGC AGCTGACTTA AGGCCAGTCA   60

AAGG   64

```
(2) INFORMATION FOR SEQ ID NO:61:

     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 54 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: single
         (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:
```

CAAAACCTTG ACATCACTTC CATTGCCAGC AACGGCGGAA GAGTGCAATG CATG   54

```
(2) INFORMATION FOR SEQ ID NO:62:

     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 46 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: single
         (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:
```

CATTGCACTC TTCCGCCGTT GCTGGCAATG GAAGTGATGT CAAGGT   46

```
(2) INFORMATION FOR SEQ ID NO:63:
```

-223-

```
        (i) SEQUENCE CHARACTERISTICS:
              (A) LENGTH: 37 base pairs
              (B) TYPE: nucleic acid
              (C) STRANDEDNESS: single
              (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:
```

CATGGCTTCC TCAGTTCTTT CCTCTGCAGC AGTTGCC   37

```
    (2) INFORMATION FOR SEQ ID NO:64:

        (i) SEQUENCE CHARACTERISTICS:
              (A) LENGTH: 37 base pairs
              (B) TYPE: nucleic acid
              (C) STRANDEDNESS: single
              (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:
```

GGGTGGCAAC TGCTGCAGAG GAAAGAACTG AGGAAGC   37

```
    (2) INFORMATION FOR SEQ ID NO:65:

        (i) SEQUENCE CHARACTERISTICS:
              (A) LENGTH: 34 base pairs
              (B) TYPE: nucleic acid
              (C) STRANDEDNESS: single
              (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:
```

ACCCGCAGCA ATGTTGCTCA AGCTAACATG GTGG   34

```
    (2) INFORMATION FOR SEQ ID NO:66:

        (i) SEQUENCE CHARACTERISTICS:
              (A) LENGTH: 32 base pairs
              (B) TYPE: nucleic acid
              (C) STRANDEDNESS: single
              (D) TOPOLOGY: linear
```

-224-

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

CGCCACCATG TTAGCTTGAG CAACATTGCT GC   32

(2) INFORMATION FOR SEQ ID NO:67:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 50 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CGCCTTTCAC TGGCCTTAAG TCAGCTGCCT CATTCCCTGT TTCAAGGAAG   50

(2) INFORMATION FOR SEQ ID NO:68:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 52 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TTTGCTTCCT TGAAACAGGG AATGAGGCAG CTGACTTAAG GCCAGTGAAA GG   52

(2) INFORMATION FOR SEQ ID NO:69:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 54 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CAAAACCTTG ACATCACTTC CATTGCCAGC AACGGCGGAA GAGTGCAATG CATG   54

-225-

(2) INFORMATION FOR SEQ ID NO:70:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 46 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CATTGCACTC TTCCGCCGTT GCTGGCAATG GAAGTGATGT CAAGGT   46

(2) INFORMATION FOR SEQ ID NO:71:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 71 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CATGGCTTCC TCAGTTCTTT CCTCTGCAGC AGTTGCCACC CGCAGCAATG TTGCTCAAGC   60

TAACATGGTG G   71

(2) INFORMATION FOR SEQ ID NO:72:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 69 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CGCCACCATG TTAGCTTGAG CAACATTGCT GCGGGTGGCA CTGCTGCAG AGGAAAGAAC   60

TGAGGAAGC   69

-226-

(2) INFORMATION FOR SEQ ID NO:73:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 104 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CGCCTTTCAC TGGCCTTAAG TCAGCTGCCT CATTCCCTGT TTCAAGGAAG CAAAACCTTG   60

ACATCACTTC CATTGCCAGC AACGGCGGAA GAGTGCAATG CATG   104

(2) INFORMATION FOR SEQ ID NO:74:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 98 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CATTGCACTC TCCGCCGTT GCTGGCAATG GAAGTGATGT CAAGGTTTTG CTTCCTTGAA   60

ACAGGGAATG AGGCAGCTGA CTTAAGGCCA GTGAAAGG   98

(2) INFORMATION FOR SEQ ID NO:75

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 36 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CCTGCAGGCA TCCAACCATG GCGTAATCAT GGTCAT 36

-227-

(2) INFORMATION FOR SEQ ID NO:76:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 35 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

AGAGCGTATC GTGAGGGATC TGATTTTAGT CGGCG 35

(2) INFORMATION FOR SEQ ID NO:77:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 35 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GCGCGGATCC ATGGGGGATC TGATTTTAGT CGGCG 35

(2) INFORMATION FOR SEQ ID NO:78:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 32 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GCGGCGCATG CGGGATCTGA TTTTAGTCGG CG 32

(2) INFORMATION FOR SEQ ID NO:79:

-228-

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 30 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

CATCGGATCC TGTCAGGAAA ATGGTTCAGC 30

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 36 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

TTAAACTATT TAGTACCATG GCGGTGCGCG CTCCTG 36

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 36 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

CAGGAGCGCG CACCGCTATG CTAGTAAATA GTTTAA 36

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 6 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear

-229-

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Met Leu Val Asn Ser Leu
1   5

(2) INFORMATION FOR SEQ ID NO:83:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 36 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA (genomic)

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

CAGGAGCGCG CACCGCCATG GTACTAAATA GTTTAA   36

(2) INFORMATION FOR SEQ ID NO:84:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Met Val Leu Asn Ser Leu
1   5

(2) INFORMATION FOR SEQ ID NO:85:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA (genomic)

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

-230-

TTACAACAAA TATAAAAACA ATGTCTTTAT   30

(2) INFORMATION FOR SEQ ID NO:86:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 30 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

TTACAACAGA TCTAAAAACA ATGTCTTTAT   30

(2) INFORMATION FOR SEQ ID NO:87:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 30 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

CTTTATGAGG GTAACATGAA TTCAAGAAGG   30

(2) INFORMATION FOR SEQ ID NO:88:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 30 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

CCTTCTTGAA TTGATGTTAC CCTCATAAAG   30

-231-

(2) INFORMATION FOR SEQ ID NO:89:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 30 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

CCTTCTTGAA ATCATGTTAC CCTCATAAAG   30

(2) INFORMATION FOR SEQ ID NO:90:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 69 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

AGCTTCGAAG AACGAAGGAA GGAGCACAGA CTTAGATTGG TATATATACG CATATTGCGG   60

CCGCGGTAC   69

(2) INFORMATION FOR SEQ ID NO:91:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 61 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

AGCTTCTTGC TTCCTTCCTG GTGTCTGAAT CTAACCATAT ATATGCGTAT AACGCCGGCG   60

C   61

-232-

(2) INFORMATION FOR SEQ ID NO:92:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

ATACGCCATG AGCCATTTTG CCATTGTGGC 30

(2) INFORMATION FOR SEQ ID NO:93:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

ATACCATATG AGCCATTTTG CCATTGTGGC 30

(2) INFORMATION FOR SEQ ID NO:94:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GCCCCCCGGG AGTCAGATCG TCTTCATGGA 30

(2) INFORMATION FOR SEQ ID NO:95:

-233-

```
        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 30 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

GCCCCCCGGG AGTCAGATCG TCTTCATGGA   30
(2) INFORMATION FOR SEQ ID NO:96:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 10 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CCGGGAATTC   10




(2) INFORMATION FOR SEQ ID NO:97:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 1200 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)



       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

ATGAGCCATT TTGCCATTGT GGCACCGCCG CTCTACAGTC ATGCGGTGGC GCTGCATGCC    60

CTGGCGCTGG AGATGGCCCA ACGCGGCCAC CGGGTGACCT TTCTCACCGG CAACGTCGCC   120

TCGCTGGCAG AGCAGGAAAC GGAGCGGGTG GCGTTCTATC CACTTCCCGC CAGCGTGCAA   180

CAGGCCCAGC GCAACGTCCA GCAGCAGAGT AACGGCAACC TGCTGCGGCT GATTGCGGCC   240

ATGTCATCCC TGACCGATGT GCTCTGCCAG CAGTTGCCCG CTATTCTACA GCGGCTGGCG   300

GTGGACGCGC TGATTGTCGA TGAGATGGAG CCCGCCGGAA GCCTGGTCGC CGAGGCGCTG   360
```

GGACTACCAT TTATCTCTAT TGCCTGCGCG CTGCCGGTCA ACCGCGAGCT GCCGCTGCCG 420

GTGATGCCGT TTCACTACGC CGAGGATAAG AGAGCCCGTG CGCGTTTTCA GGTCAGCGAA 480

CGGATCTACG ATGCGCTGAT GTACCCGCAC GGGCAGACGA TCCTGCGCCA CGCCCAGCGC 540

TTTGGTTTGC CGGAGCGCAG GCGTCTCGAC GAGTGTCTCT CGCCCCTGGC GCAGATTAGC 600

CAGTCCGTTC CGGCCCTCGA CTTCCCACGC CGGGCGCTGC CGAACTGTTT TACCTACGTG 660

GGAGCACTGC GCTATCAGCC CCCGCCGCAG GTAGAACGCT CGCCACGCAG CACGCCGCGG 720

ATCTTTGCCT CGCTGGGCAC CCTCCAGGGC CACCGTCTAC GCCTGTTTCA GAAGATCGCC 780

CGCGCCTGTG CCAGCGTGGG CGCGGAGGTG ACCATTGCCC ACTGCGATGG CCTGACGCCC 840

GCCCAGGCCG ACTCGCTCTA CTGCGGCGCG ACGGAGGTGG TCAGCTTTGT CGACCAGCCG 900

CGCTACGTTG CCGAGGCTAA TCTGGTGATC ACCCACGGCG GTCTCAATAC CGTACTGGAT 960

C_GCTGGCTG CCGCGACGCC GGTGCTGGCG GTGCCACTCT CTTTCGACCA GCCCGCCGTG 1020

GCTGCCCGGC TGGTCTATAA CGGGCTGGGT CGCCGGGTAT CGCGCTTTGC CAGACAGCAG 1080

ACGCTGGCGG ATGAGATTGC CCAACTGCTG GGGGATGAGA CGCTGCATCA GCGTGTGGCG 1140

ACGGCCCGCC AGCAGCTTAA CGACGCCGGG GGCACGCCCC GTTGCGGCGA CCCTGATTGA 1200

(2) INFORMATION FOR SEQ ID NO:98:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 59 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

    Met Ala Ser Ser Val Leu Ser Ser Ala Ala Val Ala Thr Arg Ser Asn
    1   5   10   15

    Val Ala Gln Ala Asn Met Val Ala Pro Phe Thr Gly Leu Lys Ser Ala
                    20   25   30

    Ala Ser Phe Pro Val Ser Arg Lys Gln Asn Leu Asp Ile Thr Ser Ile
                    35   40   45

    Ala Ser Asn Gly Gly Arg Val Gln Cys Met Gln
    50   55

-235-

(2) INFORMATION FOR SEQ ID NO:99:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 489 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

```
Met Glu Lys Thr Val Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
1               5               10              15

Ala Ile Arg Leu Gln Ala Ala Gly Ile Pro Thr Val Leu Leu Glu Gln
        20              25              30

Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Trp His Asp Gln Gly Phe
        35              40              45

Thr Phe Asp Ala Gly Pro Thr Val Ile Thr Asp Pro Thr Ala Leu Glu
        50              55              60

Ala Leu Phe Thr Leu Ala Gly Arg Arg Met Glu Asp Tyr Val Arg Leu
65              70              75              80

Leu Pro Val Lys Pro Phe Tyr Arg Leu Cys Trp Glu Ser Gly Lys Thr
                85              90              95

Leu Asp Tyr Ala Asn Asp Ser Phe Glu Leu Glu Ala Gln Ile Thr Gln
            100             105             110

Phe Asn Pro Arg Asp Val Glu Gly Tyr Arg Arg Phe Leu Ala Tyr Ser
        115             120             125

Gln Ala Val Phe Gln Glu Gly Tyr Leu Arg Leu Gly Ser Val Pro Phe
    130             135             140

Leu Ser Phe Arg Asp Met Leu Arg Ala Gly Pro Gln Leu Leu Lys Leu
145             150             155             160

Gln Ala Trp Gln Ser Val Tyr Gln Ser Val Ser Arg Phe Ile Glu Asp
            165             170             175

Glu His Leu Arg Gln Ala Phe Ser Phe His Ser Leu Leu Val Gly Gly
        180             185             190

Asn Pro Phe Thr Thr Ser Ser Ile Tyr Thr Leu Ile His Ala Leu Glu
        195             200             205

Arg Glu Trp Gly Val Trp Phe Pro Glu Gly Gly Thr Gly Ala Leu Val
    210             215             220
```

-236-

```
Asn Gly Met Val Lys Leu Phe Thr Asp Leu Gly Gly Glu Ile Glu Leu
225             230             235                 240

Asn Ala Arg Val Glu Glu Leu Val Val Ala Asp Asn Arg Val Ser Gln
                245             250                 255

Val Arg Leu Ala Asp Gly Arg Ile Phe Asp Thr Asp Ala Val Ala Ser
            260             265             270

Asn Ala Asp Val Val Asn Thr Tyr Lys Lys Leu Leu Gly Thr Ile Pro
            275             280             285

Val Gly Gln Lys Arg Ala Ala Arg Leu Glu Arg Lys Ser Met Ser Asn
    290             295             300

Ser Leu Phe Val Leu Tyr Phe Gly Leu Asn Gln Pro His Ser Gln Leu
305             310             315                 320

Ala His His Thr Ile Cys Phe Gly Pro Arg Tyr Arg Glu Leu Ile Asp
            325             330             335

Glu Ile Phe Thr Gly Ser Ala Leu Ala Asp Asp Phe Ser Leu Tyr Leu
            340             345             350

His Ser Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Pro Pro Cys Ala
        355             360             365

Ser Phe Tyr Val Leu Ala Pro Val Pro His Leu Gly Asn Ala Pro Leu
    370             375             380

Asp Trp Ala Gln Glu Gly Pro Lys Leu Arg Asp Arg Ile Phe Asp Tyr
385             390             395                 400

Leu Glu Glu Arg Tyr Met Pro Gly Leu Arg Ser Gln Leu Val Thr Gln
            405             410             415

Arg Ile Phe Thr Arg Gln Thr Ser Arg His Ala Trp Ile Ala Ile Leu
            420             425             430

Gly Ser Leu Phe Ile Glu Pro Pro Ser Leu Thr Gln Gly Leu Phe Ala
        435             440             445

Ala Asn Ala Thr Arg His Ser Asn Leu Tyr Leu Val Ala Ala Gly Thr
    450             455             460

His Pro Gly Ala Gly Ile Pro Gly Val Val Gly Leu Ala Glu Ser Thr
465             470             475                 480

Ala Ser Leu Met Ile Glu Asp Leu Gln
            485
```

(2) INFORMATION FOR SEQ ID NO:100:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

ATAAAGACAT TGTTTTTAGA TCTGTTGTAA   30

-238-

WHAT IS CLAIMED IS:

1.   An isolated DNA segment comprising a
nucleotide sequence that contains at least 850 base
pairs that define a structural gene for the <u>Erwinia</u>
5   <u>herbicola</u> enzyme geranylgeranyl pyrophosphate synthase
(E.C.2.5.1.29) and DNA variants thereof encoding an
enzyme exhibiting substantially the same biological
activity.

10   2.   The isolated DNA segment as described in
claim 1 wherein said structural gene is selected from
the group consisting of those present within the
approximately 1030 base pair Nco I-Eco RV restriction
fragment of plasmid pARC417BH having the ATCC accession
15   No. 40755; the approximately 1000 base pair Nco I-Pvu II
restriction fragment of plasmid pARC498D having the ATCC
accession No. 40757, and the approximately 1150 base
pair Nco I-Pvu II restriction fragment of plasmid
pARC489B having ATCC accession No. 40758.

20

           3.   A recombinant DNA molecule comprising a
vector operatively linked to an exogenous DNA segment
that contains at least 850 base pairs defining a
structural gene for the <u>Erwinia</u> <u>herbicola</u> enzyme
25   geranylgeranyl pyrophosphate synthase (E.C. 2.5.1.29)
and DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity, and a
promoter suitable for driving the expression of said
enzyme in a compatible host organism.

30

           4.   The recombinant DNA molecule as described
in claim 3 wherein said host organism is a prokaryote.

           5.   The recombinant DNA molecule as described
35   .   in claim 3 wherein said host organism is a higher plant.

-239-

6.   A method for preparing the _Erwinia_ _herbicola_ enzyme geranylgeranyl pyrophosphate synthase, comprising the steps of:

a.   initiating a culture, in a nutrient medium, of prokaryotic or eukaryotic host cells transformed with a recombinant DNA molecule comprising an expression vector compatible with said cells operatively linked to an exogenous DNA segment containing at least 850 base pairs defining the _Erwinia_ _herbicola_ structural gene for geranylgeranyl pyrophosphate synthase, DNA variants and analogs thereof encoding an enzyme exhibiting substantially the same biological activity; and

b.   maintaining said culture for a time period sufficient for said cells to express said geranylgeranyl pyrophosphate synthase protein molecule.

7.   The DNA segment as defined in claim 6 wherein said structural gene is present within the approximately 1000 base pair Nco I-Pvu II restriction endonuclease fragment of plasmid pARC489D having the ATCC accession No. 40757.

8.   An isolated DNA segment comprising a nucleotide sequence that contains at least 927 base pairs defining an _Erwinia_ _herbicola_ structural gene for phytoene synthase and DNA variants thereof encoding an enzyme exhibiting substantially the same biological activity.

9.   The isolated DNA segment as described in claim 10 wherein said structural gene is present within the approximately 1238 base pair Pvu II-Eco RI restriction endonuclease fragment of plasmid pARC140N.

5

10

15

20

25

30

35

-240-

10.   A recombinant DNA molecule comprising a
vector operatively linked to an exogenous DNA segment
that contains at least 927 base pairs defining a
structural gene for the _Erwinia herbicola_ enzyme

5   phytoene synthase, and DNA variants thereof encoding an
enzyme exhibiting substantially the same biological
activity, and a promoter suitable for driving the
expression of said enzyme in a compatible host organism.

10   11.   The recombinant DNA molecule as described
in claim 10 wherein said host organism is a prokaryote.
coli.

12.   The recombinant DNA molecule as described
15   in claim 10 wherein said host organism is a higher
plant.

13.   A method for preparing the enzyme
phytoene synthase, comprising the steps of:

20   a. initiating a culture, in a nutrient
medium, of prokaryotic or eukaryotic host cells
transformed with an exogenous recombinant DNA molecule
comprising an expression vector compatible with said
cells operatively linked to a DNA segment containing at

25   least 927 base pairs defining the structural gene for
Erwinia herbicola phytoene synthase and DNA variants
thereof encoding an enzyme exhibiting substantially the
same biological activity; and

                b.   maintaining said culture for a time
30   period sufficient for said cells to express said
phytoene synthase protein molecule.

-241-

14.   A method for producing phytoene, comprising the steps of:

a.   initiating a culture, in a nutrient medium, of prokaryotic or eukaryotic host cells transformed with a recombinant DNA molecule containing an expression system that comprises one or more expression vectors compatible with said cells operatively linked to an exogenous DNA segment comprising (i) a nucleotide base sequence that contains at least 850 base pairs defining a structural gene for Erwinia herbicola geranylgeranyl pyrophosphate synthase, and DNA variants thereof encoding an enzyme exhibiting substantially the same biological activity, and (ii) a nucleotide base sequence that contains at least 927 base pairs defining an Erwinia herbicola structural gene for phytoene synthase, and DNA variants thereof encoding an enzyme exhibiting substantially the same biologically activity; and

b.   maintaining said culture for a time period sufficient for said cells to express phytoene.

15.   The method of preparation described in claim 14 wherein said transformed host is a prokaryote.

16.   The method of preparation described in claim 14 wherein said transformed host is a higher plant.

17.   A plasmid vector containing a DNA segment that encodes GGPP synthase selected from the group consisting of plasmid pARC417BH having ATCC accession number 40755, plasmid pARC489B having ATCC accession number 40758, plasmid pARC489D having ATCC accession number 40757 and plasmid pARC145G having ATCC accession number 40753.

-242-

18.   A plasmid vector containing a DNA
segment that encodes phytoene synthase selected from the
group consisting of plasmid pARC285 having ATCC
accession number 40756, plasmid pARC140N having ATCC
5   accession number 40754, and plasmid pARC145G having ATCC
accession number 40753.

19.   An isolated DNA segment comprising a
nucleotide sequence of at least 1470 base pairs that
10   defines a structural gene for the <u>Erwinia herbicola</u>
enzyme phytoene dehydrogenase-4H and DNA variants
thereof encoding an enzyme exhibiting substantially the
same biological activity.

15   20.   The isolated DNA segment as described in
claim 19 wherein said structural gene is selected from
the group consisting of those present within the
approximately 1505 base pair Nco I-Nco I restriction
fragment of plasmid pARC496A having the ATCC accession
20   No. 40803, the approximately 1508 base pair Sal I-Sal I
restriction fragment of plasmid pARC146D having ATCC
Accession No. 40801, the approximately 1506 base pair
Sph I-Nco I restriction fragment of plasmid pATC228
having ATCC accession No. 40802, and the approximately
25   2450 base pair Xba I-Xba I restriction fragment of
plasmid pATC1616 having ATCC accession No. 40806.

21.   A recombinant DNA molecule comprising a
vector operatively linked to an exogenous DNA segment
30   that contains a nucleotide sequence of at least about
1470 base pairs defining a structural gene for the
<u>Erwinia herbicola</u> enzyme phytoene dehydrogenase-4H and
DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity, and a

-243-

promoter suitable for driving the expression of said
enzyme in a compatible host organism.

22.   The recombinant DNA molecule as described
in claim 21 wherein said host organism is a prokaryote
E. coli.

23.   The recombinant DNA molecule as described
in claim 21 wherein said host organism is a higher
plant.

24.   A method for preparing the Erwinia
herbicola enzyme phytoene dehydrogenase-4H, comprising
the steps of:
        a.   initiating a culture, in a nutrient
medium, of prokaryotic or eukaryotic host cells
transformed with a recombinant DNA molecule comprising
an expression vector compatible with said cells
operatively linked to an exogenous DNA segment
containing a nucleotide sequence of at least about 1470
base pairs defining the Erwinia herbicola structural
gene for phytoene dehydrogenase-4H and DNA variants
thereof encoding an enzyme exhibiting substantially the
same biological activity; and
                b.   maintaining said culture for a time
period sufficient for said cells to express said
phytoene dehydrogenase-4H protein molecule.

25.   The method as described in claim 24
wherein said host cells are eukaryotic cells of a higher
plant.

26.   A method for producing lycopene
comprising the steps of:

-244-

a.   initiating a culture, in a nutrient
medium, of prokaryotic or eukaryotic host cells that
provide phytoene, and are transformed with a recombinant
DNA molecule containing an expression system that

5   comprises an expression vector compatible with said host
cells operatively linked to an exogenous DNA segment
comprising a nucleotide base sequence of at least about
1470 base pairs defining an <u>Erwinia</u> <u>herbicola</u> structural
gene for phytoene dehydrogenase-4H and DNA variants

10   thereof encoding an enzyme exhibiting substantially the
same biologically activity; and

b.   maintaining said culture for a time
period sufficient for said cells to express phytoene
dehydrogenase-4H and for said expressed phytoene

15   dehydrogenase-4H to convert the provided phytoene into
lycopene.

27.   The method of preparation described in
claim 26 wherein said transformed host is a prokaryote.

20

28.   The method of preparation described in
claim 26 wherein said transformed host is a higher
plant.

25   29.   The method of preparation described in
claim 25 wherein phytoene is provided to said host cells
through an expression system comprising one or more
expression vectors compatible with said host cells
operatively linked to an exogenous DNA segment

30   comprising (i) a nucleotide base sequence that contains
at least 850 base pairs defining a structural gene for
<u>Erwinia</u> <u>herbicola</u> geranylgeranyl pyrophosphate synthase
and DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity and (ii) a

35   nucleotide base sequence that contains at least 927 base

-245-

pairs defining an <u>Erwinia</u> <u>herbicola</u> structural gene for phytoene synthase and DNA variants encoding an enzyme thereof exhibiting substantially the same biologically activity.

5

      30.   The method of preparation described in claim 29 wherein said host cells are <u>E. coli</u>, said expression vector for phytoene dehydrogenase-4H is pARC496A having ATCC accession No. 40803, and host

10 <u>E. coli</u> cells are also transformed with plasmid pARC489D having ATCC accession No. 40757 and plasmid pARC140N having ATCC accession No. 40759, the expression of said plasmids pARC489D and pARC140N providing phytoene to said host cells.

15

      31.   The method of preparation described in claim 29 wherein said host cells are <u>S. cerevisiae</u>, said expression vector for phytoene dehydrogenase-4H is pARC146D having ATCC accession No. 40801, and said host

20 <u>S. cerevisiae</u> cells are also transformed with plasmid pARC145G having ATCC accession No. 40753, the expression of said plasmid pARC145G providing phytoene to said host cells.

25   32.   The method of preparation described in claim 29 wherein said host cells are <u>R. sphaeroides</u> and said expression vector for phytoene dehydrogenase-4H is pATC228 having ATCC accession No. 40802.

30   33.   A plasmid vector containing a DNA segment that encodes phytoene dehydrogenase-4H selected from the group consisting of plasmid pARC496A having ATCC accession No. 40803, plasmid pARC146D having ATCC accession No. 40801, plasmid pATC228 having ATCC

accession No. 40802, and plasmid pATC1616 having ATCC
accession No. 40806.

34.   An isolated DNA segment comprising a
nucleotide sequence of at least about 1,125 base pairs
that define a structural gene for the _Erwinia herbicola_
enzyme lycopene cyclase and DNA variants thereof
encoding an enzyme exhibiting substantially the same
biological activity.

35.   The isolated DNA segment as described in
claim 34 wherein said structural gene is selected from
the group consisting of those present within the
approximately 1142 base pair Sph I-Bam HI DNA fragment
of sequence base pairs contained in plasmid pARC1509
having the ATCC accession number 40850 and the
approximately 1156 base pair Hind III to Bam HI
restriction fragment of plasmid pARC1509 having ATCC
accession number 40850.

36.   A recombinant DNA molecule comprising a
vector operatively linked to an exogenous DNA segment
that contains a nucleotide sequence of at least about
1125 base pairs defining a structural gene for the
Erwinia herbicola enzyme lycopene cyclase and DNA
variants thereof encoding an enzyme exhibiting
substantially the same biological activity, and a
promoter suitable for driving the expression of said
enzyme in a compatible host organism.

37.   The recombinant DNA molecule as described
in claim 36 wherein said host organism is a prokaryote.

38.   The recombinant DNA molecule as described
in claim 36 wherein said host organism is a higher
plant.

5   39.   A method for preparing the _Erwinia_
herbicola enzyme lycopene cyclase, comprising the steps
of:
            a.   initiating a culture, in a nutrient
medium, of procaryotic or eukaryotic host cells
10   transformed with a recombinant DNA molecule comprising
an expression vector compatible with said cells
operatively linked to an exogenous DNA segment
containing a nucleotide sequence of at least about 1125
base pairs defining the _Erwinia_ _herbicola_ structural
15   gene for lycopene cyclase and DNA variants thereof
encoding an enzyme exhibiting substantially the same
biological activity; and
            b.   maintaining said culture for a time
period sufficient for said cells to express said
20   lycopene cyclase protein molecule.

            40.   A method for producing beta-carotene
comprising the steps of:
            a.   initiating a culture, in a nutrient
25   medium, of procaryotic or eukaryotic host cells that
provide lycopene, and are transformed with a recombinant
DNA molecule containing an expression system that
comprises an expression vector compatible with said host
cells operatively linked to an exogenous DNA segment
30   comprising a nucleotide base sequence of at least about
1125 base pairs defining an _Erwinia_ _herbicola_ structural
gene for lycopene cyclase and DNA variants thereof
encoding an enzyme exhibiting substantially the same
biological activity; and

-248-

b.   maintaining said culture for a time period sufficient for said cells to expr ss lycopene cyclase and for said expressed lycopene cyclase to convert the provided lycopene into beta-carotene.

41.   The method of preparation described in claim 40 wherein said transformed host is a prokaryote.

42.   The method of preparation described in claim 40 wherein said transformed host is a higher plant.

43.   The method of preparation described in claim 40 including the further step of recovering said beta-carotene from said maintained culture.

44.   The method of preparation described in claim 40 wherein lycopene is provided to said host cells through an expression system comprising one or more expression vectors compatible with said host cells operatively linked to an exogenous DNA segment comprising:

(i) a nucleotide base sequence that contains at least 850 base pairs defining a structural gene for <u>Erwinia herbicola</u> geranylgeranyl pyrophosphate synthase, and DNA variants thereof encoding an enzyme exhibiting substantially the same biological activity,

(ii) a nucleotide base sequence that contains at least 927 base pairs defining an <u>Erwinia herbicola</u> structural gene for phytoene synthase and DNA variants thereof encoding an enzyme exhibiting the same biological activity, and,

(iii) a nucleotide base sequence that contains at least 1470 base pairs defining an <u>Erwinia herbicola</u> structural gene for phytoene dehydrogenase-4H

-249-

and DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity.

5   45.   The method of preparation described in
claim 44 wherein said host cells are E. coli, said
expression vector for lycopene cyclase is pARC1510
having ATCC accession number 40851, and said host
E. coli cells are also transformed with plasmid pARC489D
having ATCC accession number 40757, plasmid pARC140N
10   having ATCC accession number 40759, and plasmid pARC496A
having ATCC accession number 40803, the expression of
said plasmids pARC489D, pARC140N, and pARC496A providing
lycopene to said host cells.

15   46.   The method of preparation described in
claim 44 wherein said host cells are S. cerevisiae and
said expression vector for lycopene cyclase is pARC1520
having ATCC accession number 40852.

20   47.   A plasmid vector containing a DNA segment
that encodes lycopene cyclase selected from the group
consisting of plasmid pARC1510 having ATCC accession
number 40851, plasmid pARC1520 having ATCC accession
number 40852, and plasmid pARC1509 having ATCC accession
25   number 40850.

           48.   An isolated DNA segment comprising a
nucleotide sequence of at least 531 base pairs that
defines a structural gene for the Erwinia herbicola
30   enzyme beta-carotene hydroxylase and DNA variants
thereof encoding an enzyme exhibiting substantially the
same biological activity.

           49.   The isolated DNA segment as described in
35   claim 48 wherein said structural gene is present within
the approximately 870 base pair Nco I-Sma I DNA fragment

-250-

of sequence base pairs contained in plasmid pARC406BH
having the ATCC accession number 40945.

50.   A recombinant DNA molecule comprising a
vector operatively linked to an exogenous DNA segment
that contains a nucleotide sequence of at least about
531 base pairs defining a structural gene for the
Erwinia herbicola enzyme beta-carotene hydroxylase and
DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity, and a
promoter suitable for driving the expression of said
enzyme in a compatible host organism.

51.   The recombinant DNA molecule as described
in claim 50 wherein said host organism is a prokaryote.

52.   The recombinant DNA molecule as described
in claim 50 wherein said host organism is a higher
plant.

53.   A method for preparing the Erwinia
herbicola enzyme beta-carotene hydroxylase comprising
the steps of:
        a.   initiating a culture, in a nutrient
medium, of prokaryotic or eukaryotic host cells
transformed with a recombinant DNA molecule comprising
an expression vector compatible with said cells
operatively linked to an exogenous DNA segment
containing a nucleotide sequence of at least 531 base
pairs defining the Erwinia herbicola structural gene for
beta-carotene hydroxylase and DNA variants thereof
encoding an enzyme exhibiting substantially the same
biological activity; and

-251-

    b.   maintaining said culture for a time
period sufficient for said cells to express said beta-
carotene hydroxylase protein molecule.

        54.   A method for producing zeaxanthin
comprising the steps of:
        a.   initiating a culture, in a nutrient
medium, of prokaryotic or eukaryotic host cells that
provide beta-carotene and are transformed with a
recombinant DNA molecule containing an expression system
that comprises an expression vector compatible with said
host cells operatively linked to an exogenous DNA
segment comprising a nucleotide base sequence of at
least 531 base pairs defining an Erwinia herbicola
structural gene for beta-carotene hydroxylase and DNA
variants thereof encoding an enzyme exhibiting
substantially the same biological activity; and
        b.   maintaining said culture for a time
period sufficient for said cells to express beta-
carotene hydroxylase and for said expressed beta-
carotene hydroxylase to convert the provided beta-
carotene into zeaxanthin.

        55.   The method of preparation described in
claim 54 wherein said transformed host is a prokaryote.

        56.   The method of preparation described in
claim 54 wherein said transformed host is a higher
plant.

        57.   The method of preparation described in
claim 54 wherein beta-carotene is provided to said host
cells through an expression system comprising one or
more expression vectors compatible with said host cells

-252-

operatively linked to an exogenous DNA segment
comprising:

        (i)   a nucleotide base sequence that
contains at least 850 base pairs defining a structural
gene for _Erwinia herbicola_ geranylgeranyl pyrophosphate
synthase and DNA variants thereof encoding an enzyme
exhibiting substantially the same biological activity,

        (ii)   a nucleotide base sequence that
contains at least 927 base pairs defining an _Erwinia
herbicola_ structural gene for phytoene synthase and DNA
variants thereof encoding an enzyme exhibiting the same
biological activity,

        (iii) a nucleotide base sequence that
contains at least 1470 base pairs defining an _Erwinia
herbicola_ structural gene for phytoene dehydrogenase-4H
and DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity, and

        (iv)   a nucleotide base sequence that
contains at least 1,125 base pairs defining an _Erwinia_
structural gene for lycopene cyclase and DMA variants
thereof encoding an enzyme exhibiting substantially the
same biological activity.

      58.   The method of preparation described in
claim 57 wherein said host cells are _E. coli_, said
expression vector for beta-carotene hydroxylase is
pARC406BH having ATCC accession number 40946, and said
host _E. coli_ cells are also transformed with plasmid
pARC489D having ATCC accession number 40757, plasmid
pARC140N having ATCC accession number 40759, plasmid
pARC496A having ATCC accession number 40803 and plasmid
pARC1509 having ATCC acession number 40850, the
expression of said plasmids pARC489D, pARC140N, pARC496A
and pARC1509 providing beta-carotene to said host cells.

-253-

59.   The method of pr paration described in
claim 57 wherein said host cells are <u>S. cerevisiae</u>, said
expression vector for beta-carotene hydroxylase is
pARC145H having ATCC accession number 40944.

5

60.   A plasmid vector containing a DNA segment
that encodes beta-carotene hydroxylase selected from the
group consisting of plasmid pARC406BH having ATCC
accession number 40945, plasmid pARC145H having ATCC
10   accession number 40944, and plasmid pARC404BH having
ATCC accession number 40943.

61.   An isolated DNA segment comprising a
nucleotide sequence of at least 1200 base pairs that
defines a structural gene for the <u>Erwinia</u> <u>herbicola</u>
15   enzyme zeaxanthin glycosylase and DNA variants thereof
encoding an enzyme exhibiting substantially the same
biological activity.

20   62.   The isolated DNA segment as described in
claim 61 wherein said structural gene is present within
the approximately 1390 base pair Nde I-Ava I DNA
fragment of sequence base pairs contained in plasmid
pARC2019 having the ATCC accession number 40974.

25

63.   A recombinant DNA molecule comprising a
vector operatively linked to an exogenous DNA segment
that contains a nucleotide sequence of at least about
1200 base pairs defining a structural gene for the
30   <u>Erwinia</u> <u>herbicola</u> enzyme zeaxanthin glycosylase and DNA
variants thereof encoding an enzyme exhibiting
substantially the same biological activity, and a
promoter suitable for driving the expression of said
enzyme in a compatible host organism.

35

-254-

64.   The recombinant DNA molecule as described in claim 63 wherein said host organism is a prokaryote.

65.   The recombinant DNA molecule as described in claim 63 wherein said host organism is a higher plant.

66.   A method for preparing the _Erwinia herbicola_ enzyme zeaxanthin glycosylase comprising the steps of:
    a.   initiating a culture, in a nutrient medium, of procaryotic or eukaryotic host cells transformed with a recombinant DNA molecule comprising an expression vector compatible with said cells operatively linked to an exogenous DNA segment containing a nucleotide sequence of at least 1200 base pairs defining the _Erwinia herbicola_ structural gene for zeaxanthin glycosylase and DNA variants thereof encoding an enzyme exhibiting substantially the same biological activity; and
    b.   maintaining said culture for a time period sufficient for said cells to express said zeaxanthin glycosylase protein molecule.

67.   A method for producing zeaxanthin diglucoside comprising the steps of:
    a.   initiating a culture, in a nutrient medium, of prokaryotic or eukaryotic host cells that provide zeaxanthin and are transformed with a recombinant DNA molecule containing an expression system that comprises an expression vector compatible with said host cells operatively linked to an exogenous DNA segment comprising a nucleotide base sequence of at least 1200 base pairs defining an _Erwinia herbicola_ structural gene for zeaxanthin glycosylase and DNA

variants thereof encoding an enzyme exhibiting
substantially the same biological activity;

        b. maintaining said culture for a time
period sufficient for said cells to express zeaxanthin
glycosylase and for said expressed zeaxanthin
glycosylase to convert the provided zeaxanthin into
zeaxanthin diglucoside; and

        c. recovering said zeaxanthin
diglucoside.

        68.   The method of preparation described in
claim 67 wherein said transformed host is a prokaryote.

        69.   The method of preparation described in
claim 67 wherein said transformed host is a higher
plant.

        70.   The method of preparation described in
claim 67 wherein zeaxanthin is provided to said host
cells through an expression system comprising one or
more expression vectors compatible with said host cells
operatively linked to an exogenous DNA segment
comprising:

        (i) a nucleotide base sequence that
contains at least 850 base pairs defining a structural
gene for _Erwinia herbicola_ geranylgeranyl pyrophosphate
synthase and DNA variants thereof encoding an enzyme
exhibiting substantially the same biological activity,

        (ii) a nucleotide base sequence that
contains at least 927 base pairs defining an _Erwinia
herbicola_ structural gene for phytoene synthase and DNA
variants thereof encoding an enzyme exhibiting the same
biological activity,

        (iii) a nucleotide base sequence that
contains at least 1470 base pairs defining an _Erwinia_

-256-

herbicola structural gene for phytoene dehydrogenase-4H
and DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity,

        (iv)   a nucleotide base sequence that
5   contains at least 1125 base pairs defining an Erwinia
herbicola structural gene for lycopene cyclase and DNA
variants thereof encoding an enzyme exhibiting
substantially the same biological activity, and

        (v)   a nucleotide base sequence that
10   contains at least 531 base pairs defining an Erwinia
herbicola structural gene for beta-carotene hydroxylase
and DNA variants thereof encoding an enzyme exhibiting
substantially the same biological activity.

15   71.   The method of preparation described in
claim 70 wherein said host cells are E. coli, said
expression vector for zeaxanthin glycosylase is pARC2019
having ATCC accession number 40974, and said host
E. coli cells are transformed with plasmid pARC489D
20   having ATCC accession number 40757, plasmid pARC140N
having ATCC accession number 40759, plasmid pARC496A
having ATCC accession number 40803, plasmid pARC1509
having ATCC accession number 40850 and plasmid pARC406BH
having ATCC accession number 40946, the expression of
25   said plasmids pARC489D, pARC140N, pARC496A, pARC1509 and
pARC406BH providing zeaxanthin to said host cells.

        72.   The method of preparation described in
claim 70 wherein said host cells are S. cerevisiae.

30

        73.   A plasmid vector containing a DNA segment
that encodes zeaxanthin glycosylase that is plasmid
pARC2019 having ATCC accession number 40974.

-257-

74.   A method for protecting a higher plant
from the herbicide norflurazon that comprises the steps
of:

        a.   transforming a higher plant to be

5   protected with a recombinant DNA molecule that encodes a
structural gene for the Erwinia herbicola enzyme
phytoene dehydrogenase-4H or a DNA variant thereof that
encode enzyme exhibiting substantially the same
biological activity;

10   b.   maintaining the transformed plant
for a time period sufficient for said phytoene
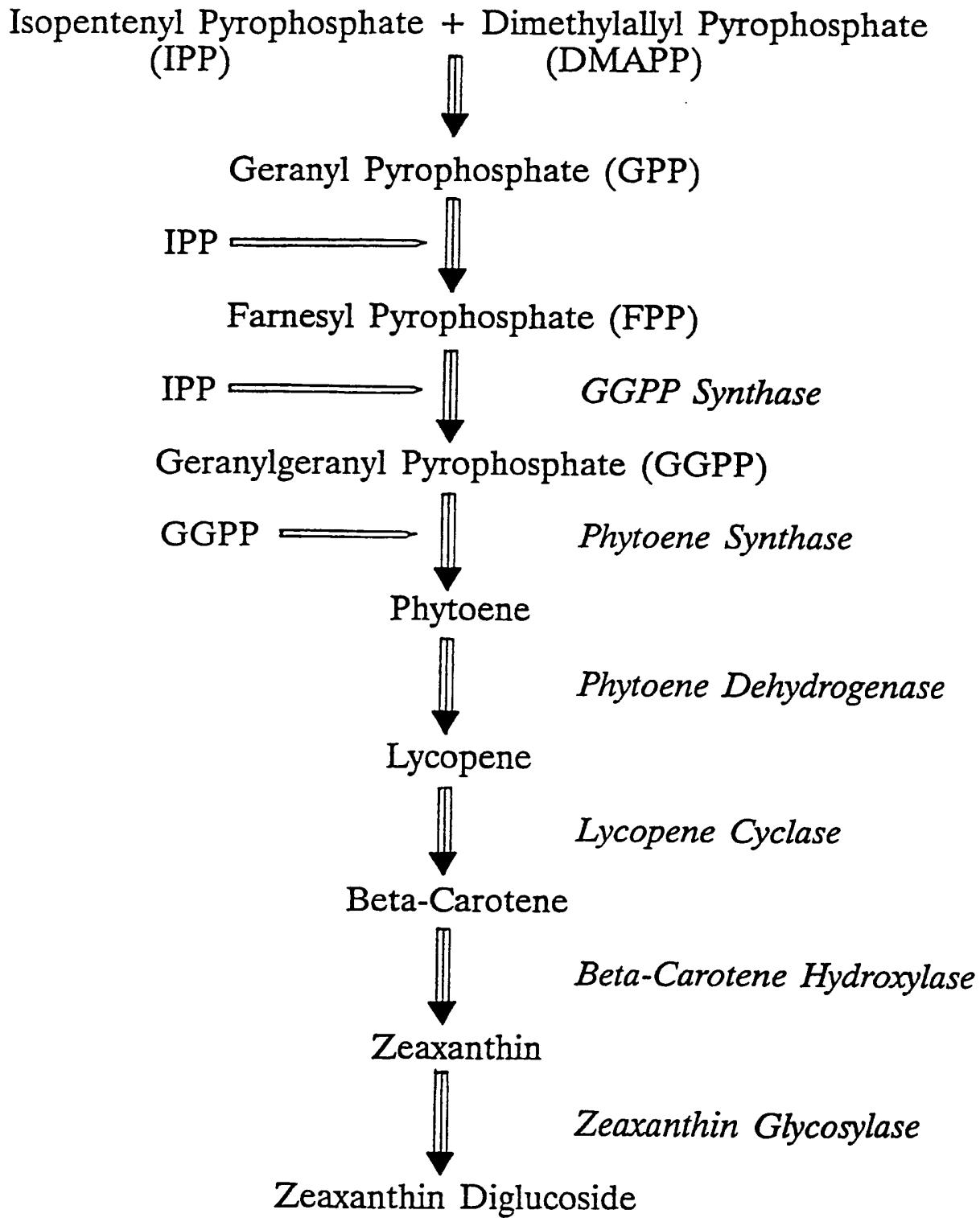dehyrogenase-4H to be expressed; and

        c.   treating the transformed plant with
a herbicidal amount of norflurazon.

15

        75.   The method as described in claim 74
wherein said recombinant DNA molecule includes an about
177 base pair sequence that encodes a chloroplast
transit peptide of tobacco ribulose bis-phosphate

20   carboxylase-oxygenase operatively linked in frame to the
5' end phytoene dehydrogenase-4H structural gene.

        76.   The method as described in claim 75
wherein said recombinant DNA molecule includes the

25   approximately 2450 base pair Xba I-Xba I fragment
present in plasmid pATC1616 having ATCC accession number
40806.

Figure 1

Carotenoid Biosynthesis Scheme

Isopentenyl Pyrophosphate + Dimethylallyl Pyrophosphate
(IPP) (DMAPP)

⇓

Geranyl Pyrophosphate (GPP)

IPP ⇒ ⇓

Farnesyl Pyrophosphate (FPP)

IPP ⇒ ⇓ *GGPP Synthase*

Geranylgeranyl Pyrophosphate (GGPP)

GGPP ⇒ ⇓ *Phytoene Synthase*

Phytoene

⇓ *Phytoene Dehydrogenase*

Lycopene

⇓ *Lycopene Cyclase*

Beta-Carotene

⇓ *Beta-Carotene Hydroxylase*

Zeaxanthin

⇓ *Zeaxanthin Glycosylase*

Zeaxanthin Diglucoside

Bgl II
AGATCTAAAGGCACAGCGTCTCATGCTTCGCACAATGTAAAACTGCTTCAGAACCTGGCG   60

Hpa I
AGAGCTATCCCGCGCCGGTCTACGGTTAACTGATACTAAAAGACAATTCAGCGGGTAACCTT   120

Nru I
GCAATGGTGAGTGGCAGTAAAGCGGGGGCGTTTCGCCCTCATCGCGAAATAGAGGTAATGAGA   180
MetValSerGlySerLysAlaGlyValSerProHisArgGluIleGluValMetArg

CAATCCATTGACGATCACCTGGCCTGGCCTGTTACCTGAAACCGACAGCCAGGATATCGTC   240
GlnSerIleAspAspHisLeuAlaGlyLeuLeuProGluThrAspSerGlnAspIleVal

Bam HI
AGCCTTGCGATGCGTGAAGGCGTCATGGCACCCGGTAAACGGATCCGCTGCTGATG   300
SerLeuAlaMetArgGluGlyValMetAlaProGlyLysArgIleArgProLeuLeuMet

CTGCTGGCCGCCCCGCGACCTCCGCTACCAGGCAGTATGCCTACGCTGTGCGATCTCGCC   360
LeuLeuAlaAlaAlaArgAspLeuArgTyrGlnGlySerMetProThrLeuLeuAspLeuAla

TGCGCCGTTGAACTGACACCATACCGGTCGCTGATGCTCGACGACACATGCCCTGCATGGAC   420
CysAlaValGluLeuThrHisThrAlaSerLeuMetLeuAspAspMetProCysMetAsp

FIGURE 2-1

3

ACCGGCCGAGCTGGCGCGGTCAGCCCACTACCCACCACAAAAAATTTGGTGAGAGCGTGGCG   480
AsnAlaGluLeuArgArgGlyGlnProThrThrHisLysLysPheGlyGluSerValAla

ATCCCTTGCCTCCGTTGGGCTGCTCTCTAAAGCCTTTGGTCTGATCGCCGCCACCGGCGAT   540
IleLeuLeuAlaSerValGlyLeuLeuIleAlaAlaPheGlyLeuLeuIleAlaAlaThrGlyAsp

CTGCCGGGGGAGAGGCCGTGCCCAGGCGGTCAACGAGCTCTACCGCCGTGGGCTGCAG   600
LeuProGlyGluArgArgAlaGlnAlaValAlaValGlyLeuGln

GGCCTGGTACTGGGGCAGTTTCGCGATCTTAACGATGCCCGCTACCCCCTGAC   660
GlyLeuValLeuGlnPheArgAspLeuAsnAspAlaAlaLeuAspArgThrProAsp

GCTATCCTCAGCACCACCACCCTCAAGACACCGGCCATTCTGTTCAGCGGCGATGCTGCAGATC   720
AlaIleLeuSerThrAsnHisLeuLysThrGlyIleLeuPheSerAlaMetLeuGlnIleIle

GTCGGCCATTGCTTCCGCCTCCGTCGCCGAGCACGCGGAGAGACGCTGCACCGCCTTCGCCCTC   780
ValAlaIleAlaSerSerAlaSerSerProSerThrArgGluThrLeuHisAlaPheAlaLeu

GACTTCGGCCAGGCGGTTCAACTGCTGGACGACGATCTGCCTGAAGAAGAACCGGT   840
AspPheGlyGlnAlaPheGlnLeuLeuAspAspLeuArgLeuArgAspAspHisProGluThrGly

FIGURE 2-2

```
AAAGATCGCAATAAGGACGCGGGAAAAATCGACGCTGGTCAACCGGCTGGGCGGCAGACGCG  900
          LysAspArgAsnLysAspAlaGlyLysSerThrLeuValAsnArgLeuGlyAlaAspAla

GCCCGGCAAAAGCTGCGCGAGCATATTGATTCCGCCGACAAACACCTCACTTTGCCTGT  960
          AlaArgGlnLysLeuArgGluHisIleAspSerAlaAspLysHisLeuThrPheAlaCys

                              Bal I

CCGCAGGGCGGCGGCCATCCGACAGTTTATGCATCTGTGGTTGGCCATCACCTTGCCGAC  1020
          ProGlnGlyGlyGlyAlaIleArgGlnPheMetHisLeuTrpPheGlyHisHisLeuAlaAsp

TGGTCACCGGTCATGAAAAATCGCCCGATACCGCCGTTTGGGTTCAAGCAGTACATAACG  1080
          TrpSerProValMetLysIleAla

ATGGAACCACATTACAGGAGTAGTGATGAATGAAGGACGAGCGCCTGTTCAGCGGTAAGA  1140

          Eco RV
          ACGATCATCTGGATATC
```

FIGURE 2-3

Bgl II
AGATCTAAAGGCACAGGCGTCTCATGCTTCGCACAATGTAAAACTGCTTCAGAACCTGGCG   60

       Hpa I
AGAGCTATCCGCGGTCTACGGGTTAACTGATACTAAAAGACAATTCAGCGGGTAACCTT   120

       Nco I   Nru I
GCAATGGGTGAGTGGCAGTAAAGCGGGGCGTCATGGCCGGAATTCGAAATAGAGGTAATGAGA   180
                       MetAlaGluPheGluIleGluValMetArg

CAATCCATTGACGATCACCTGGCCTGGCCCTGTTACCTGAAACCGACAGCCAGGATATCGTC   240
GlnSerIleAspAspHisLeuAlaAlaGlyLeuLeuProGluThrAspSerGlnAspIleVal

                     Bam HI
AGCCTTGCGATGCCTGAAGGCGTCATGGCCACCCGGTAAACGGATCCGTCCGCTGCTGATG   300
SerLeuAlaMetArgGluGlyValMetAlaProGlyLysArgIleArgProLeuLeuMet

CTGCTGGCCCCGCCGACCTCCGCTACCAGGCAGTATGCCTACGCTGCTCGATCTCGCC   360
LeuLeuAlaAlaArgAspLeuArgTyrGlnGlySerMetProThrLeuLeuAspLeuAla

TGCGCCGTTGAACTGACCCCATACCGGCCTGATGCTGCGACGACATGCCCTGCATGGAC   420
CysAlaValGluLeuThrHisThrAlaSerLeuMetLeuAspAspMetProCysMetAsp

FIGURE 3-1

ACCGGCGAGCTGCGCGGGTCAGCCACTACCCACAAAAATTGGTGAGAGCGTGGCG   480
AsnAlaGluLeuArgArgGlyGlnProThrThrHisLysLysPheGlyGluSerValAla

ATCCTTGCCTCCGTTGGGCTGCTCTCTAAAGCCTTTGGTCTGATCGCGCCACGGCGAT   540
IleLeuAlaSerValGlyLeuLeuSerLysAlaPheGlyLeuIleAlaAlaThrGlyAsp

CTGCCCGGGGGAGAGGCGGTGCCCAGGCCGGTCAACGAGCTCTCTACCGGCGGGCTGCAG   600
LeuProGlyGluArgArgAlaGlnAlaValAlaAsnGluLeuSerThrAlaValGlyLeuGln

GGCCTGGTACTGGGGCAGTTCGCGATCTTAACGATGCCCGCCCTCGACCCCCTGAC   660
GlyLeuValLeuGlyGlnPheArgAspLeuAsnAspAlaAlaLeuAspArgThrProAsp

GCTATCCTCCAGCACCAACCACCTCAAGACCGGCATTCTGTTCAGCGCGATGCTGCAGATC 720
AlaIleLeuSerThrAsnHisLeuLysThrGlyIleLeuPheSerAlaMetLeuGlnIle

GTCGGCCATTGCTTCCGCCTCGTCGCCGAGCACGCGGAGAGACGCTGCACGCCCTTCGCCCTC 780
ValAlaIleAlaSerProSerThrArgGluThrLeuHisAlaPheAlaLeu

GACTTCGGCCAGGCGGTTCAACTGCTGGACGATCTGCGTGACGATCACCCGGAAACCGGT   840
AspPheGlyGlnAlaPheGlnLeuLeuAspAspLeuArgAspAspHisProGluThrGly

FIGURE 3-2

```
AAAGATCGCAATAAGGACGCGGGAAAATCGACGCTGGTCAACCGGCTGGGGCGCAGACGCG 900
LysAspArgAsnLysAspAlaGlyLyLysSerThrLeuValAsnArgLeuGlyAlaAspAla

GCCCGGCAAAAGCTGCGCGAGCACATATTGATTCCGCCGACAAACACCTCACTTTGCCTGT 960
AlaArgGlnLysLeuArgGluHisIleAspSerAlaAspLysHisLeuThrPheAlaCys
                           Bal I

CCGCAGGGCGGCGCCATCCGACAGTTTATGCATCTGTGGTTGGCCATCACCTTGCCGAC 1020
ProGlnGlyGlyAlaIleArgGlnPheMetHisLeuTrpPheGlyHisHisLeuAlaAsp

TGGTCACCGGTCATGAAAATCGCCTGATACCGCCCTTTGGGTTCAAGCAGTACATAACG 1080
TrpSerProValMetLysIleAla

ATGGAACCACATTACAGGAGTAGTGATGAATGAAGGACGAGGCGCCTTGTTCAGCGGTAAGA 1140

Eco RV
ACGATCATCTGGATATC
```

FIGURE 3-3

FIGURE 4-1

```
           Bgl II                                    Nco I                                          60
GATTGAGGATCTGCAATGAGCCAACCGCCCGCTGCTTGACCACCGCCAGACCATGGCC
                  MetSerGlnProProLeuLeuAspHisAlaThrGlnThrMetAla

                                                                                                  120
AACGGCTCGAAAGTTTGCCACCGCTGCGAAGCTGTTCGACCCGGCCACCCGCCGTAGC
AsnGlySerLysSerPheAlaThrAlaAlaLysLeuPheAspProAlaThrArgArgSer

                                                                                                  180
GTGCTGATGCTCTACACCTGGTGCCGCCACTGCGATGACGTCATTGACGACCAGACCCAC
ValLeuMetLeuTyrThrTrpCysArgHisCysAspAspValIleAspAspGlnThrHis

                                                                                                  240
GGCTTCGCCAGCGAGGCCGGCGAGGAGGCCACCCAGCGCCTGGCCCGGCTGCGC
GlyPheAlaSerGluAlaAlaAlaGluGluAlaThrGlnArgLeuAlaArgLeuArg

                                        Bam HI                                                     300
ACGCTGACCCTGGCCGGTTTGAAGGGCCGAGATGCCAGGATCCGGCCTTCGCTGCCTTT
ThrLeuThrLeuAlaAlaPheGluGlyAlaGluMetGlnAspProAlaPheAlaAlaPhe

                                                                                                  360
CAGGAGGTGGGCGCTGACCCCACGGTATTACGCCCGCATGGCCGCTCGATCACCTCGACGGC
GlnGluValAlaAlaLeuThrHisGlyIleIleThrProArgMetAlaLeuAspHisLeuAspGly

                                                                                                  420
TTTGCGATGGACGTGGCTCAGACCCGGTATGTCACCTTTGAGGATACGCTGCTGCTACTGC
PheAlaMetAspValAlaGlnThrArgTyrValThrPheGluAspThrLeuArgTyrCys
```

480

TATCACGTGGCGGGCGTGGTGGTGTCTGATGAGTGATGGCCGTGCGGGATGAG
TyrHisValAlaGlyValValGlyLeuMetMetAlaArgValValArgAspGlu

Sma I

540

CGGGTGCTGGGATCTCGGCCCTGCCGATCTGGGCCCTTCCAGCTGACGAATATGGCCCGG
ArgValLeuAspArgAlaCysAspLeuGlyLeuAlaPheGlnLeuThrAsnMetAlaArg

Pst I

600

GATATTATTGACGATGCGGGCTATTGACCGCTGCTATCTGCCCGCCGAGTGGCTGCAGGAT
AspIleIleAspAspAlaAlaIleAspArgCysTyrLeuProAlaGluTrpLeuGlnAsp

660

GCCGGGCTGGCCCCGGGGGAGAACTATGCCGCGGGGGAGAATCGCCCGCCGCGGGTGG
AlaGlyLeuAlaProGluAsnTyrAlaAlaArgGluAsnArgProAlaLeuAlaArgTrp

720

CGGAGGCTTATTGATGCCGGAGAGCCGTACTACATCTCCTCCCAGCCGGGCTACACGAT
ArgArgLeuIleAspAlaAlaGluProTyrTyrIleSerSerGlnAlaGlyLeuHisAsp

780

CTGCGGGCGGGCTCCGCGTGGGATCGCGCCAGCCGTCTACCGGGAGATCGGT
LeuArgArgSerAlaTrpAlaIleAlaThrAlaArgSerValTyrArgGluIleGly

840

ATTAAGGTAAAAGGCGGGGAGGCAGCCGCCTGGGATCGCGCCAGCACCAGCAAAGGT
IleLysValLysAlaAlaGlyGlySerAlaTrpAspArgAlaArgGlnHisThrSerLysGly

FIGURE 4-2

GAAAAATTGCCATGCTGATGGGCGGGCACCGGGGGCAGGTTATTCGGGCGGAAGACGACGAGG   900
GluLysIleAlaMetLeuMetAlaAlaProGlyGlnValIleArgAlaAlaLysThrThrArg

GTGACGCCGGCGTCCGGCCGGTCTTGGCAGCGTCCCGTTAGGCGGGCGGCCATGACGTT   960
ValThrProArgProAlaGlyLeuTrpGlnArgProVal
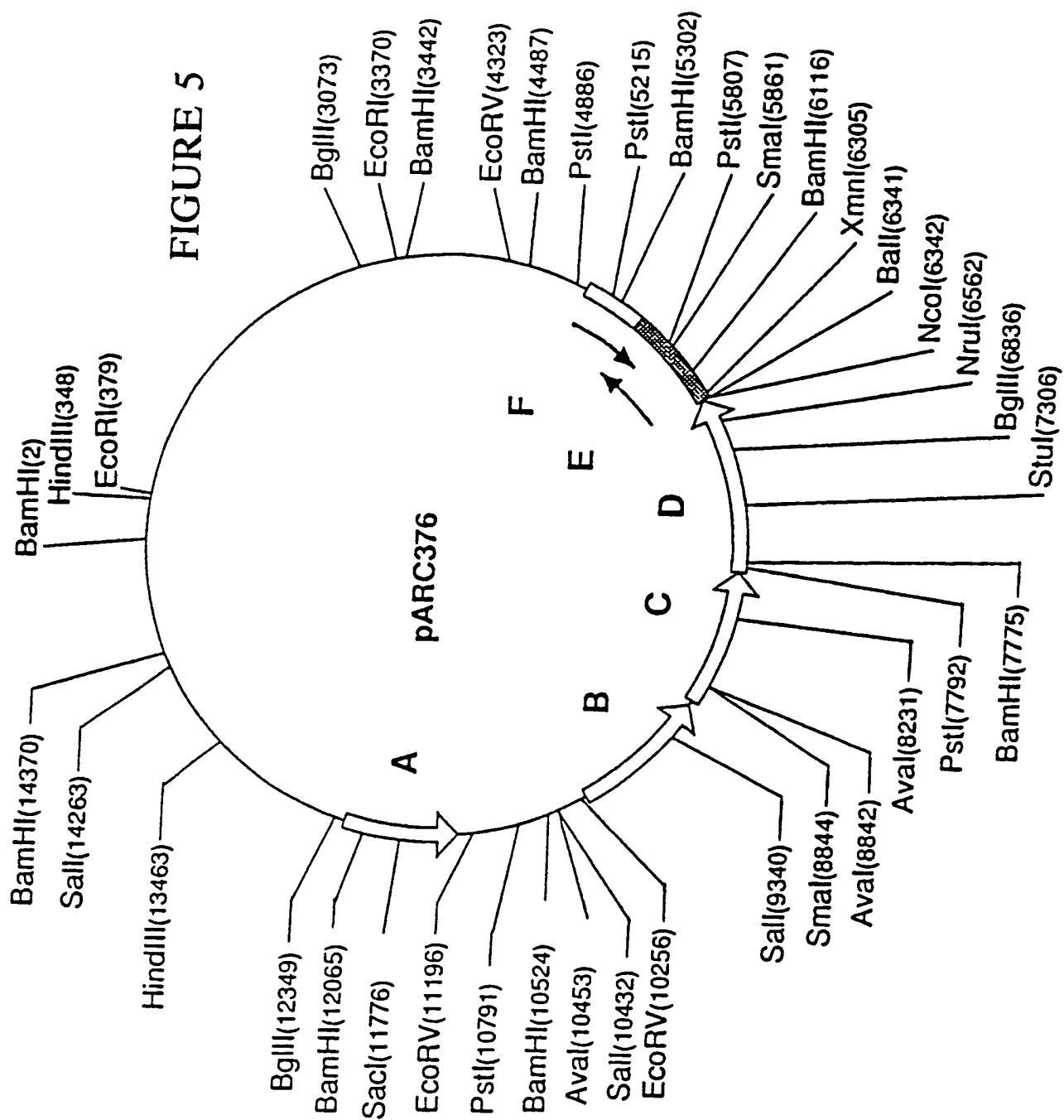
CACGCCAGGATCGCCTGTAGTCGGGCAGGCCTTGCGGGCGTAAATAAAACCGAAGGAGACGC   1020

AGCCCTCCCGGCGGCACCGCCGGTGGTGCAGCGGTGGGCGACGTAGAGCCGCTTCAGGT   1080

Bam HI
AGCCCCCGGGCGGGATCCAGTGGAAGGGCCAGCGCTGATGCACCAGACCGTCGTGCACCA   1140

Pst I
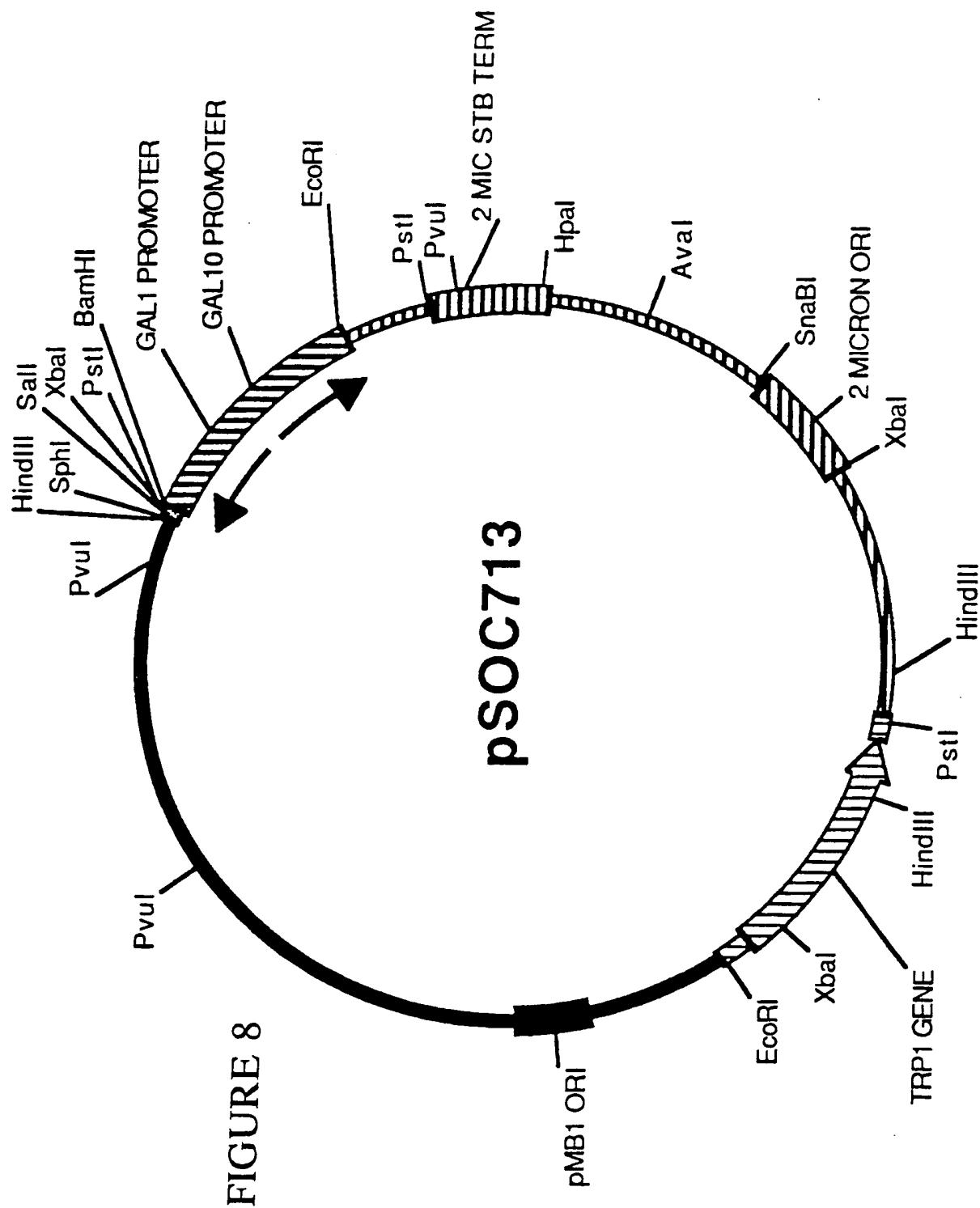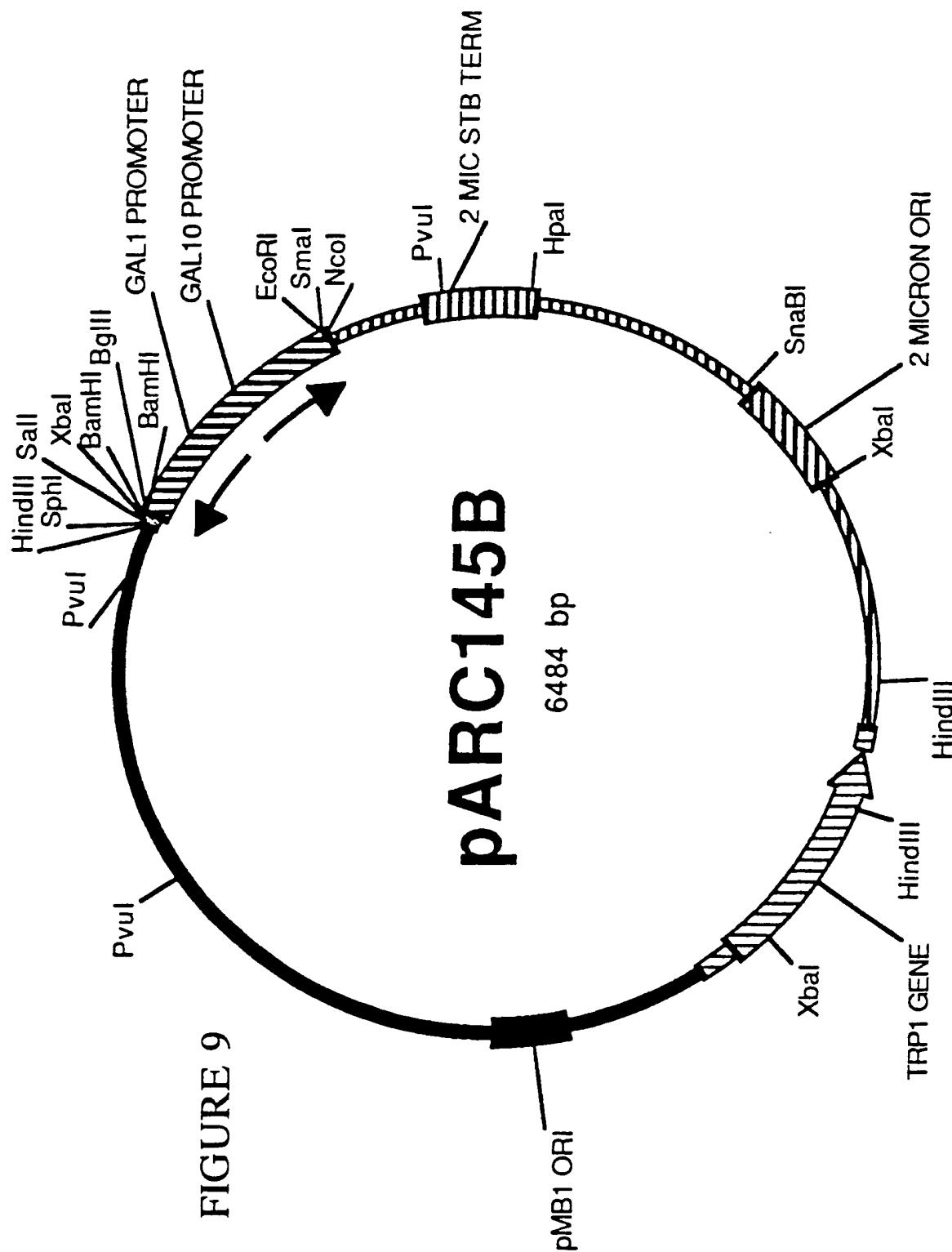GGAAGTAGAGCAGGCCATAGACCGTCATGCCGCAGCCAATCCACTGCAGGGGCCAAAC   1200

FIGURE 4-3

FIGURE 5

12

FIGURE 6

HindIII/SphI
PstI SalI
XbaI BamHI
SmaI KpnI
SstI
EcoRI
NcoI
HpaI
EcoRV
PvuII

PvuII

Rec7
Promoter

pMB1
origin

pARC306A
2519 bp

Amp - Region

DraII
AatII

SspI

XmnI

ScaI

PvuI

FspI

BglI

FIGURE 7

PUC18

pARC135

PGK PROMOTER

HindIII

BstXI

BglII

PGK TERMINATOR

EcoRI
SacI
KpnI
SmaI
BamHI
XbaI SalI PstI SphI HindIII

FIGURE 8

pSOC713

HindIII
SphI
SalI
Xbal
PstI
BamHI
GAL1 PROMOTER
GAL10 PROMOTER
EcoRI
PstI
PvuI
2 MIC STB TERM
HpaI
AvaI
SnaBI
2 MICRON ORI
Xbal
HindIII
PstI
HindIII
Xbal
TRP1 GENE
EcoRI
pMB1 ORI
PvuI
PvuI

15

FIGURE 9

16

FIGURE 10

FIGURE 11-1

Nco I
TAAACCATGGAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCGCCTGGCCGATT   60
     MetLysThrValValIleGlyAlaGlyPheGlyGlyLeuAlaLeuAlaIle

Pst I   Bam HI
CGCCTGCAGGGCGGCAGGGATCCCAACCCTACTGCTGGAGCAGCCCGGGCGGT   120
ArgLeuGlnAlaAlaAlaGlyIleProThrValLeuLeuGluGlnArgAspLysProGlyGly

CGGGCCTACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGCGGTGATCACC   180
ArgAlaTyrValTrpHisAspGlnGlyPheThrPheAspAlaAlaProThrValIleThr

GATCCTACCGCGCTTGAGGGCGCTGTTCACCCTGGCCGGCAGGGCCATGGAGGATTACGTC   240
AspProThrAlaLeuGluGlyAlaLeuPheThrLeuAlaGlyArgArgMetGluAspTyrVal

AGGCTGCTCGGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGAAGACCCTCGAC   300
ArgLeuLeuProValLysProPheTyrArgLeuCysTrpGluSerGlyLysThrLeuAsp

TATGCTAACGACAGCTTCGAGTTGAGGCGCAGATTACCCAGTTCAACCCCGCGACGTC   360
TyrAlaAsnAspSerPheGluLeuGluAlaGlnIleThrGlnPheAsnProArgAspVal

GAGGGCTACCGGCGCTTTCTGGCTTACTCCCAGGCGGGTATTCCAGGAGGGATATTGCGC   420
GluGlyTyrArgArgPheLeuAlaValPheGlnGluGlyGluTyrLeuArg

FIGURE 11-2

Nru I
CTGGGCAGCGGTGCCGTTCCTCTCTTTCGGACATGCTGCCGCGGGCCGCAGCTGCTT   480
LeuGlySerValProPheLeuSerPheArgAspMetLeuArgAlaGlyProGlnLeuLeu

AAGCTCCAGGGCTGGCAGAGCGTCTACCAGTCGGTTCGCGTTTATTGAGGATGAGCAT   540
LysLeuGlnAlaTrpGlnSerValTyrGlnSerValSerArgPheIleGluAspGluHis

CTGCGGCAGGCCTTCTGTTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCG   600
LeuArgGlnAlaPheSerPheHisSerLeuLeuValGlyGlyAsnProPheThrThrSer

TCCATCTACACCCCTGATCCACCGCGCCCCTGAGCGGGAGTGGGGGGTCTCGGTTCCCCTGAGGGC   660
SerIleTyrThrLeuIleHisAlaLeuGluArgGluTrpGlyValTrpPheProGluGlyGly

GGCACCGGGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGGGGGGAGATC   720
GlyThrGlyGlyAlaLeuValAsnGlyMetValLysLeuPheThrAspLeuGlyGlyGluIle

Sma I
GAACTCAACGCCCGGGTGGTGGCCGATAACCGGTGGCCGGAAGAGCTCGAAGAGCTCCGG   780
GluLeuAsnAlaArgValValAlaAspAsnArgValSerGlnValArg

CTCGCGGGATGGTCGGATCTTTGACACCGACGCCCTAGCCTGACGTGGTGAAC   840
LeuAlaAspGlyArgIlePheAspThrAspAlaValAlaSerAsnAlaAspValValAsn

FIGURE 11-3

```
ACCTATAAAAGCTGCTCGGCACCATACCGGTGGGGCAGAGAAGCGGGCACGGCTGGAG        900
ThrTyrLysLysLeuLeuGlyThrIleProValGlyGlnLysLysArgAlaAlaArgLeuGlu

CGCAAGAGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCC        960
ArgLysSerMetSerAsnSerLeuPheValLeuTyrPheGlyLeuAsnGlnProHisSer
                                                    Bgl II

CAGCTGGCGCACCATACCCATCTGTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATC       1020
GlnLeuAlaHisHisThrIleCysPheGlyProArgTyrArgGluLeuIleAspGluIle

TTTACCGGGCAGCGGCGGGATGACTTCTCGCTCTCACCTGCACTCGCCCTGCGTGACC        1080
PheThrGlySerAlaLeuAlaAspAspPheSerLeuTyrLeuHisSerProCysValThr

GATCCCTCGCTGCCCTCCCCGGTGCTGCCCAGCTTCTACGTGCCGCCAT               1140
AspProSerLeuAlaProProCysAlaSerPheTyrValLeuAlaProValProHis

CCTTGGCAACGGCGCTGGACTGGGGCCAGGAGGGGCCGAAGCTGCCGACCGCCATCTTT       1200
LeuGlyAsnAlaProLeuAspTrpAlaGlnGluGlyProLysLeuProThrAlaIlePhe

GACTACCTTGAAGAGGCTATATGCCCGGCCTAGCCAGCTGGTGACCCAGCCGGATC         1260
AspTyrLeuGluGluArgTyrMetProGlyLeuArgSerGlnLeuValThrGlnArgIle
```

TTTACCCGGCCAGACTTCACGACACGCTTGGATCGCGGCGATCTTGGGATCGCTTTTCATCGAG   1320
PheThrArgGlnThrSerArgHisAlaTrpIleAlaIleLeuGlySerLeuPheIleGlu

CCGCCTTCGTTGACCCAAGGCTTGTTCGCCCGCCAAACGCCGACACGACATTCAAACCTCTAC   1380
ProProSerLeuThrGlnGlyLeuPheAlaAlaAsnAlaThrArgHisSerAsnLeuTyr

CTGGTGGCCGCAGGTACTCACCCTGGGCGGGCATTCCTGGCGTAGTGGGCCTCGCCGAA   1440
LeuValAlaAlaGlyThrHisProGlyAlaGlyIleProGlyValValGlyLeuAlaGlu

AGCACCGCCAGCCTGATGATTGAGGATCTGCAATGAGCCAACCCGCCTGCTTGACCACG   1500
SerThrAlaSerLeuMetIleGluAspLeuGln

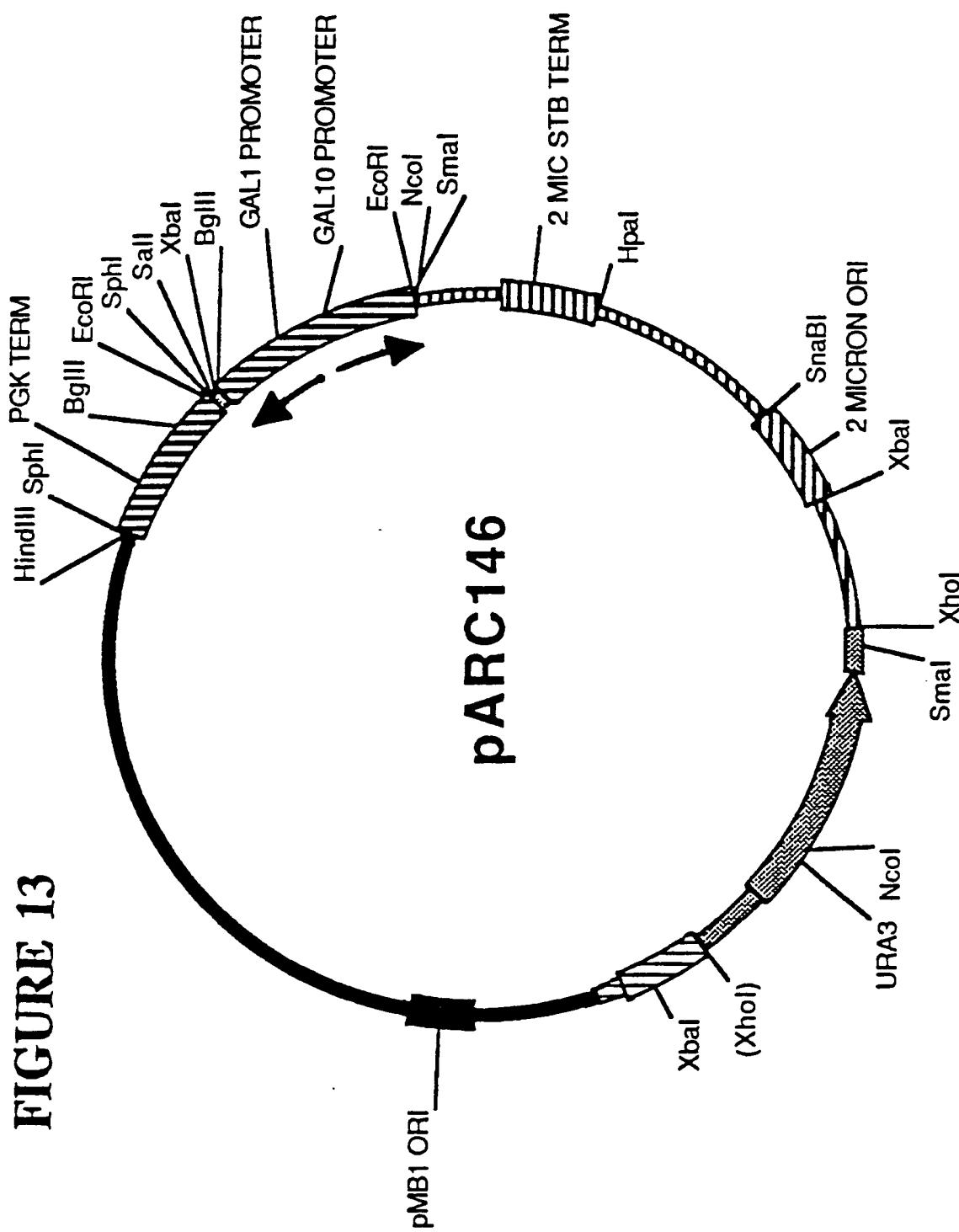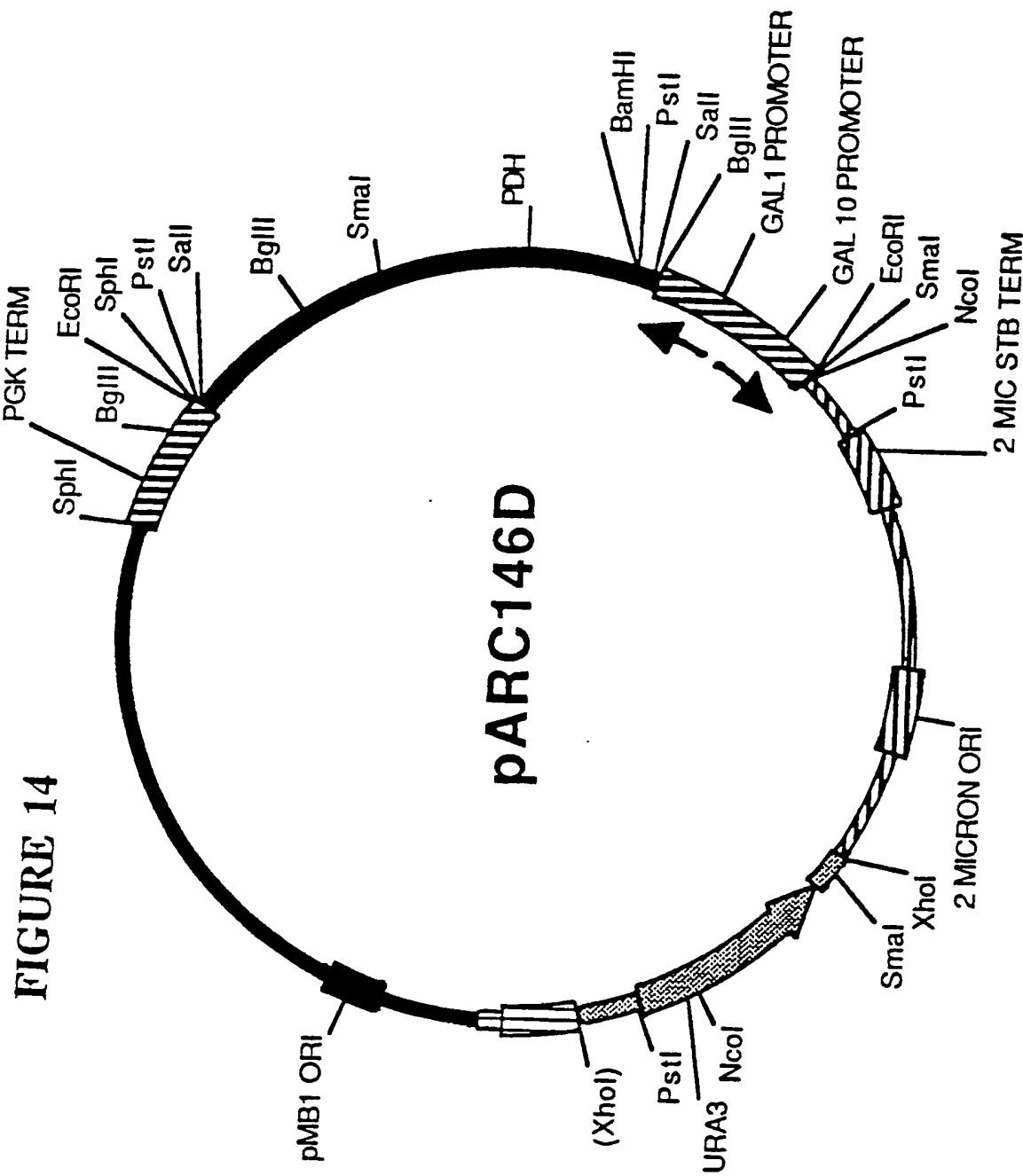   Nco I   Bal I
CCACGCAGACCATGGCCA

FIGURE 11-4

FIGURE 12

22

FIGURE 13

pARC146

HindIII SphI
PGK TERM
BgIII EcoRI
SphI
SalI
XbaI
BgIII
GAL1 PROMOTER
GAL10 PROMOTER
EcoRI
NcoI
SmaI
2 MIC STB TERM
Hpal
SnaBI
2 MICRON ORI
XbaI
XhoI
SmaI
URA3 NcoI
(XhoI)
XbaI
pMB1 ORI

23

FIGURE 14

pARC146D

GAGG

Sal I
TCGACGACGATGGAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCGGCCTGGCGATT   64
　　　　　MetGluLysThrValValIleGlyAlaGlyPheGlyGlyGlyLeuAlaIle

Pst I　　　Bam HI
CGCCTGCAGGCGGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAGCCCGGCGGT   124
ArgLeuGlnAlaAlaAlaGlyIleProThrValLeuLeuGluGlnArgAspLysProGlyGly

CGGGCCCTACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGCCCCACGGTGATCACC   184
ArgAlaTyrValTrpHisAspGlnGlyPheThrPheAspAlaGlyProThrValIleThr

GATCCTACCGCCGCTTGAGGCGTGTTCACCCTGGCCGGCCATGGAGGATTACGTC   244
AspProThrAlaLeuGluAlaLeuPheThrLeuAlaLeuAlaGlyArgArgMetGluAspTyrVal

AGGCTGCTGCCGGTAAAACCCTTCTACCGACTCTGCTGGAGTCCGGGAAGACCCTCGAC   304
ArgLeuLeuProValLysProPheTyrArgLeuCysTrpGluSerGlyLysThrLeuAsp

TATGCTAACGACAGCTTCGAGTCTTGAGGCGCGCAGATTACCCCGCGACGTC   364
TyrAlaAsnAspSerPheGluLeuGluAlaGlnIleThrGlnPheAsnProArgAspVal

FIGURE 15-1

25

```
GAGGGCTACCGGCGGCTTCTGGCTTACTCCCAGGCGGGTATTCCAGGAGGGGATATTGCGC    424
GluGlyTyrArgArgPheLeuAlaTyrSerGlnAlaValPheGluGluGlyTyrLeuArg
```

```
                              Nru I
CTCGGGCAGCGGTGCCGTTCCTCTCTTTCGCGACATGCTGCGCCGGCCAGCTGCTT    484
LeuGlySerValProPheLeuSerPheArgAspMetLeuArgAlaGlyProGlnLeuLeu
```

```
AAGCTCCAGGCGGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCAT    544
LysLeuGlnAlaTrpGlnSerValTyrGlnSerValSerArgPheIleGluAspGluHis
```

```
CTGCGCGGCAGGCCCTTCGTTCCACTCCCCTGCTAGGCGGGCAACCCCCTTCACCACCTCG    604
LeuArgGlnAlaPheSerPheHisSerLeuLeuLeuValGlyGlyAsnProPheThrThrSer
```

```
TCCATCTACACCCCTGATCCACGCCCCTTGAGCGGGAGTGGGGGTCTGGTTCCCTGAGGGC    664
SerIleTyrThrLeuIleHisAlaLeuGluArgGluTrpGlyValTrpPheProGluGly
```

```
GGCACCGGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTCGGGGGGGGAGATC    724
GlyThrGlyAlaLeuValAsnGlyMetValLysLeuPheThrAspLeuGlyGlyGlyGluIle
```

```
                  Sma I
GAACTCAACGCCCGGGTCGGTGGCCGATAACCGGTAAGCCAGGTCCGG    784
GluLeuAsnAlaArgValGluGluLeuLeuValAlaAlaAspAsnArgValSerGlnValArg
```

FIGURE 15-2

CTCGGGGATGGTCGGATCTTTGACACCGACGCCCGTAGCCTCGAACGCTGACGTGGTGAAC 844
LeuAlaAspGlyArgIlePheAspThrAspAlaValAlaSerAsnAlaAspValValAsn

ACCTATAAAAGCTGCTCGGCACCATACCGGTGGGCAGAAGCGGGCACGGCTGGAG 904
ThrTyrLysLysLeuLeuGlyThrIleProValGlyGlnLysArgAlaAlaArgLeuGlu

CGCAAGAGCATGAGCAACTCGCTGTTGTGCTCTTCGGCCTGAACCAGCCTCATTCC 964
ArgLysSerMetSerAsnSerLeuPheValLeuTyrPheGlyLeuAsnGlnProHisSer

Bgl II
CAGCTGGCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATC 1024
GlnLeuAlaHisHisThrIleCysPheGlyProArgTyrArgGluLeuIleAspGluIle

TTTACCGGGCAGCGCGCTGGCGATGACTTCTCGCTCTACCTGCACTCGCCCCTGCGTGACC 1084
PheThrArgGlnThrSerArgHisAlaTrpIleAlaIleLeuGlySerLeuPheIleGlu

GATCCCCTGCGCCTCCCCGTGCGCGCCCCAGCTTCTACGTGCTGGCCCCGGTGCCGGCCAT 1144
AspProSerLeuAlaProProCysAlaSerPheTyrValLeuAlaProValProHis

FIGURE 15-3

CTTGGCAACGGCCGGCTGGACTGGGGCGCAGGAGGGCCGAAGCTGCGGCGACCGGCATCTTT   1204
LeuGlyAsnAlaProLeuAspTrpAlaGlnGluGluGlyProLysLeuArgAspArgIlePhe

GACTACCTTGAAGAGCGCTATATGCCCGGCCTGGTGACCCAGCGGATC   1264
AspTyrLeuGluGluArgTyrMetProGlyLeuArgSerGlnleuValThrGlnArgIle

TTTACCCGGCAGACTTCACGACACGCTTGGATCGGCGATCTTGGATCGGCTTTTCATCGAG   1324
PheThrArgGlnThrSerArgHisAlaTrpIleAlaIleLeuLeuGlySerLeuPheIleGlu

CCGCCTTCGTTGACCCAAGGCTTGTTCCCGCCAAACGCGACACGACCATTCAAACCTCTAC   1384
ProProSerLeuThrGlnGlyLeuPheAlaAlaAsnAlaThrArgHisSerAsnLeuTyr

CTGGTGGCCGCCAGGTACTCACCCTGGCGGGGCCATTCCTGGCGATTCCTGGGCATTCCTGGGCATT   1444
LeuValAlaAlaGlyThrHisProGlyAlaGlyIleProGlyValValGlyLeuAlaGlu

AGCACCGGCCAGCCTGATGATTGAGGATCTGCAATGAGCCAACCGCCGGCTTGACCCACG   1504
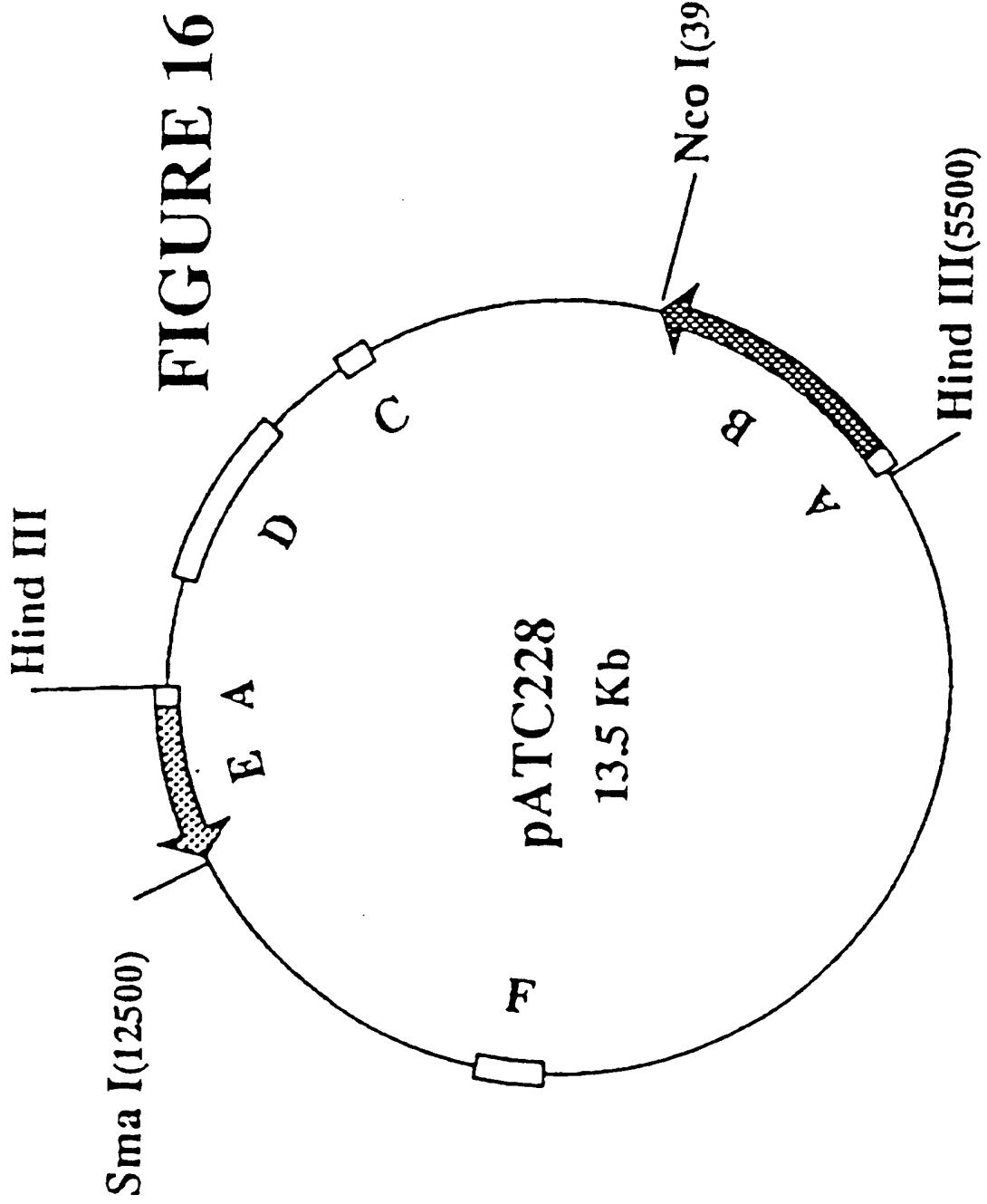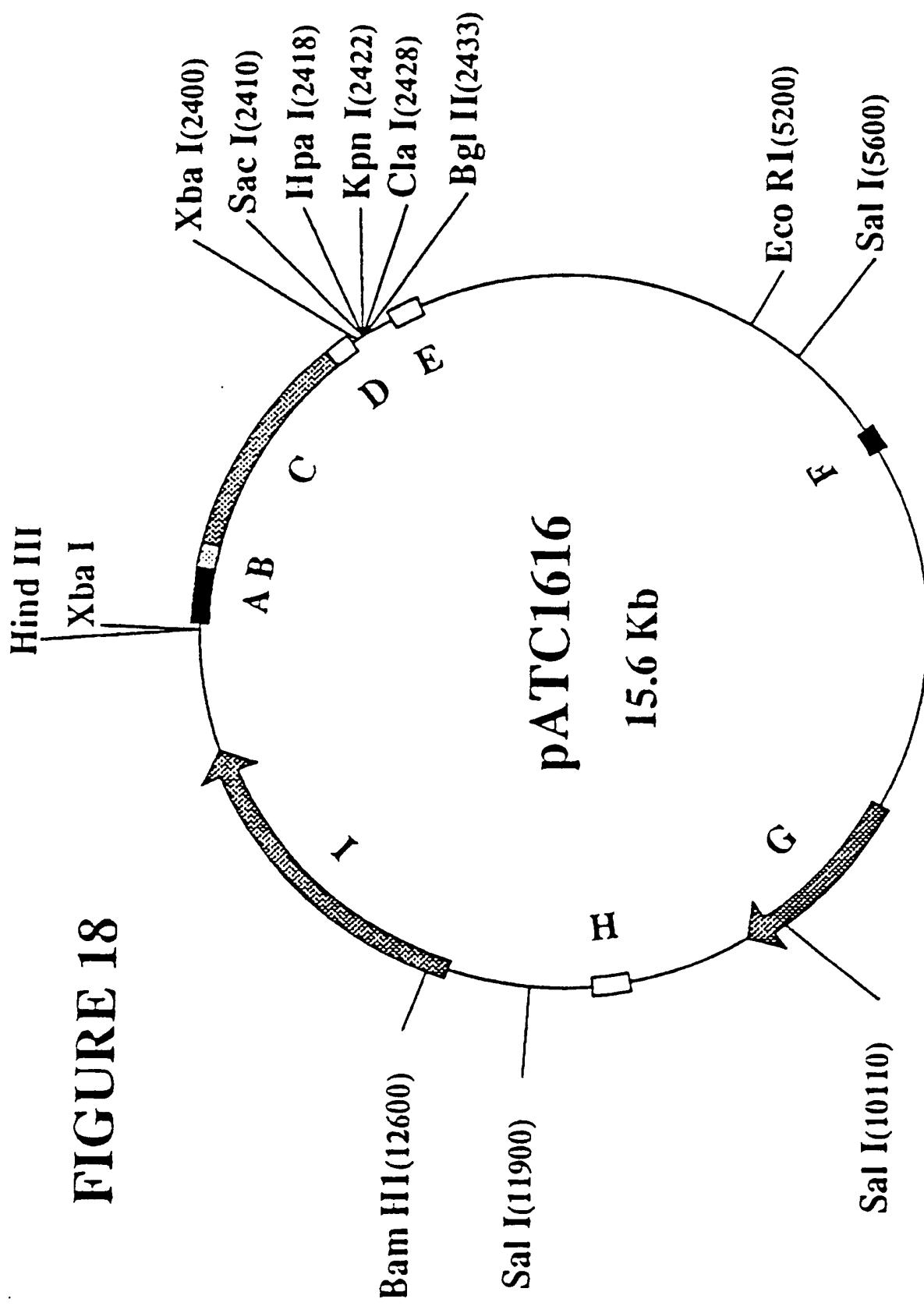SerThrAlaSerLeuMetIleGluGluAspLeuGln

Sal I
CCACGTCGACCATGGCCA

FIGURE 15-4

28

FIGURE 16

29

FIGURE 17

```
                                                          27
ATG GCT TCC TCA GTT CTT TCC TCT GCA GTT GCC ACC CGC AGC
MET Ala Ser Ser Val Leu Ser Ser Ala Val Ala Thr Arg Ser

                      *       *               81
AAT GTT GCT CAA GCT AAC ATG GTG GCG CCT TTC ACT GGC CTT AAG
Asn Val Ala Gln Ala Asn MET Val Ala Pro Phe Thr Gly Leu Lys

            108                                       135
TCA GCT GCC TCA TTC CCT GTT TCA AGG AAG CAA AAC CTT GAC ATC
Ser Ala Ala Ser Phe Pro Val Ser Arg Lys Gln Asn Leu Asp Ile

                      162
ACT TCC ATT GCC AGC AAC GGC GGA GTG AGA GTG CAA TGC ATG CAG
Thr Ser Ile Ala Ser Asn Gly Gly Arg Val Gln Cys MET Gln
```

FIGURE 18

pATC1616
15.6 Kb

Xba I(2400)
Sac I(2410)
Hpa I(2418)
Kpn I(2422)
Cla I(2428)
Bgl II(2433)

Eco R1(5200)
Sal I(5600)

Hind III
Xba I

Bam H1(12600)
Sal I(11900)
Sal I(10110)

A B C D E F G H I

Sph I

AGGGAGTGAGAGCCGTATCGTGAGGGATCTGATTTTAGTCGGCCGCGGGCCTGGCCAACGGG   60
GCGGCGCATGCGG
    MetArgAspLeuIleLeuValGlyArgGlyLeuAlaAsnGly

CTGATCGCCTGGCCTCTGCGCCAGCGCTACCCGCAGCTTAACCTGCTGCTGATCGAGGCC   120
LeuIleAlaTrpArgLeuArgGlnArgTyrProGlnLeuAsnLeuLeuLeuIleGluAla

GGGGAGCAGCCCGGGAACCATACCTGGTCATTCCATGAAGACGATCTGACTCCCGGG   180
GlyGluGlnProGlyGlyAsnHisThrTrpSerPheHisGluAspAspLeuThrProGly

CAGCACGCCTGGCCTGGCCCCGCTGGTGGCCCACGCCCTGGCCTATGAGGTGCAGTTT   240
GlnHisAlaTrpLeuAlaProLeuValAlaHisAlaTrpProGlyTyrGluValGlnPhe

CCCGATCTTCGCCGTCGCGGGCCTGCGCGGGCTACTACTCCATTACCTACGGCTTTGCC   300
ProAspLeuArgArgArgLeuAlaArgLeuLeuLeuHisTyrTyrSerIleThrSerGluArgPheAla

GAGGCCCTGCATCAGGCGCTGGGGGGAGAACATCTGGCTAAAACTGTTCGGTGAGCGAGGTG   360
GluAlaLeuHisGlnAlaLeuGluGlyGluAsnIleTrpLeuAsnCysSerValSerGluVal

TTACCCAATAGCGTGCCCTTGCCAACGGTGAGGGCGCTTGCCGGAGCCGGTGATTGAC   420
LeuProAsnSerValArgLeuLeuAlaAsnGlyGluAlaAlaGlyAlaValIleAsp

FIGURE 19-1

GGACGCGGGCGGTGACCGGCCAGTTCGGCGATGCAAAACCGGCTATCAGCTCTTTCTTGGTCAG   480
GlyArgGlyValThrAlaSerSerAlaMetGlnThrGlyTyrGlnLeuLeuPheLeuGlyGln

CAGTGGCGGCTGACACAGCCCCACGGCCTACCGGCCTGATGGATGCCACGGTG   540
GlnTrpArgLeuThrGlnProHisGlyLeuThrValProIleLeuMetAspAlaThrVal

GCGCCAGCAGCAGGGGCTATCGCTTTGTCTACACGCTGCCCGCCGACACGCTGCTG   600
AlaGlnGlnGlnGlyLeuSerAlaTyrArgPheValTyrThrLeuProLeuSerAlaAspThrLeuLeu

ATCGAGGATACGCGCTACGCCAATGTCCCGCAGCGTGATGATAATGCCCTACGCCAGACG   660
IleGluAspThrArgTyrAlaAsnValProGlnArgAspAspAsnAlaLeuArgGlnThr

GTTACCGACTATGCTCACAGCAAAGGGTGGCAGCTGCCCAGCTTGAACGCGAGGAGACC   720
ValThrAspTyrAlaHisSerLysGlyTrpGlnLeuAlaGlnLeuGluArgGluGluThr

GGCTGTCTGCCGATTACCTGGCGGGTGACATCCAGGCTCTGTGGCCGGCGT   780
GlyCysLeuProIleThrTrpArgValThrSerArgLeuCysGlyProMetArgArgArg

GCCGGCGTCGGGAATGCGGGCTATTCCACCCTACCACTGGCTATTCCGCTGCCGCTG   840
AlaAlaSerGlyMetArgAlaGlyLeuLeuPheHisProThrThrGlyTyrSerLeuProLeu

FIGURE 19-2

GCGGTGGCCCTTGCCGACGGCGATTGCCGACAGCCCGGCTGGGCAGGGTTCCGCTCTAT   900
AlaValAlaLeuAlaAspAlaIleAlaAspSerProArgLeuGlySerValProLeuTyr

CAGCTCACCCGGCCAGTTGCCGAACGCCACTGGCCAGGCAGGGATTCTTCCGCCTGCTG   960
GlnLeuThrArgGlnPheAlaGluArgHisTrpArgArgGlnGlyPhePheArgLeuLeu

AACCGGATGCTTTCCTGGCGGGCGGGGAGGAGAACCGCTGCCGGGTGATGCAGCGCTTT   1020
AsnArgMetLeuPheLeuAlaGlyArgGluGluAsnArgTrpArgValMetGlnArgPhe

TATGGGCTGCCGAGCCCACCGTAGAGCGGCTTTTACGGCCGGCTCTCTCTCTTTGAT   1080
TyrGlyLeuProGluProThrValGluProThrValGluArgPheTyrAlaGlyArgLeuSerLeuPheAsp

AAGGCCCCGCATTTGACGGGCAAGCCACCGGTTCCGCTGGGCCACCGGTTCCGCTGGGCC   1140
LysAlaArgIleLeuThrGlyLysProProValProLeuAlaLysSerGlyGlyArgArg

TGAACCATTTTCCTGACACAGACGAGATAAAAGGATGAAAAAACCGTTGTGATTGGCGCAGG   1200
<u>GATCCGATG</u>

Bam HI

Pst I

CTTTGGTGGCCTGGCGGCGATTCGCCCTGCAG

FIGURE 19-3

34

```
TGAGTCGGCAGGAGCGGCACCGGCTATGCTAGTAAATAGTTTAATCGTCATCTTGACCGTT      60

ATTGCGGATGGGAGGGCCATCGCCGCCGTTTACCCACCGCTACATTATGCACGGCTGGGGATGG     120

CGCTGGCATGAGCCACCATACCCCGCGCAAGGGGCGTATTTGAGCTAAAACGATCTCTTT     180

GCGGTGGTGTGTTGCGGGGGTGGGGCTATCGCGCTGATTGCCGTGGGCACGGGCCGGTTTGG     240

CCCCTGCAGTGGATTGGCGGGCATGACGGTCTATGGCCTGCTCTACTTCCTGGTGCAC     300

GACGGTCTGGTGCATCAGCGGCTGGCCCTTCCACTGGATCCCGGGGCTACCTGAAG     360

CGGCTCTACGTCCCACCGCCCTGCACCACCGGCGGTGCGGGGAGGGCTGCCGTCTCC     420
```

Bam HI

FIGURE 20-1

35

```
TTCGGTTTTATTTACGCCCGCAAGCCTGCCGACCTACAGGGCGATCCTGCCGGTGAACGTCAT        480

GGCCGCCCGCCTAAACGGGACGCTGCCAAAGACCGGCGGTCACCCTCGTCG                    540

TCTTCGCCCGAATAAACCTGCCCGGTGCCGCCATCAGCAGTGGCAATTTTTTTCACCTTTGC         600

TGGTGTGCTGGGGGCGATCCCAGGCGCTCCCGCCGCTTTACCTTAATACCGATCT                660

CCCGGTAGACGCTGCGGGTGCGGGTGGATCGCCCACGCCCGCCCGCAGATCGTGTA               720

GCCCGGCCTGGGAGGAGATGTAGTACGGCTTCTGCGGCATCAATAAGCCTCCCGCCACCGCG         780

CCAGCGCGGGGGCGATTCTCCCGGGCCATAGTTCTCCCGGGGCCCGGGCATCCTGCA             840
```

FIGURE 20-2

Sma I

GCCACTCGGCGGGGCAGATAGCAGCGGTCAATAGCCGCCATCGTCAATAATATCCCGGGCCA   900

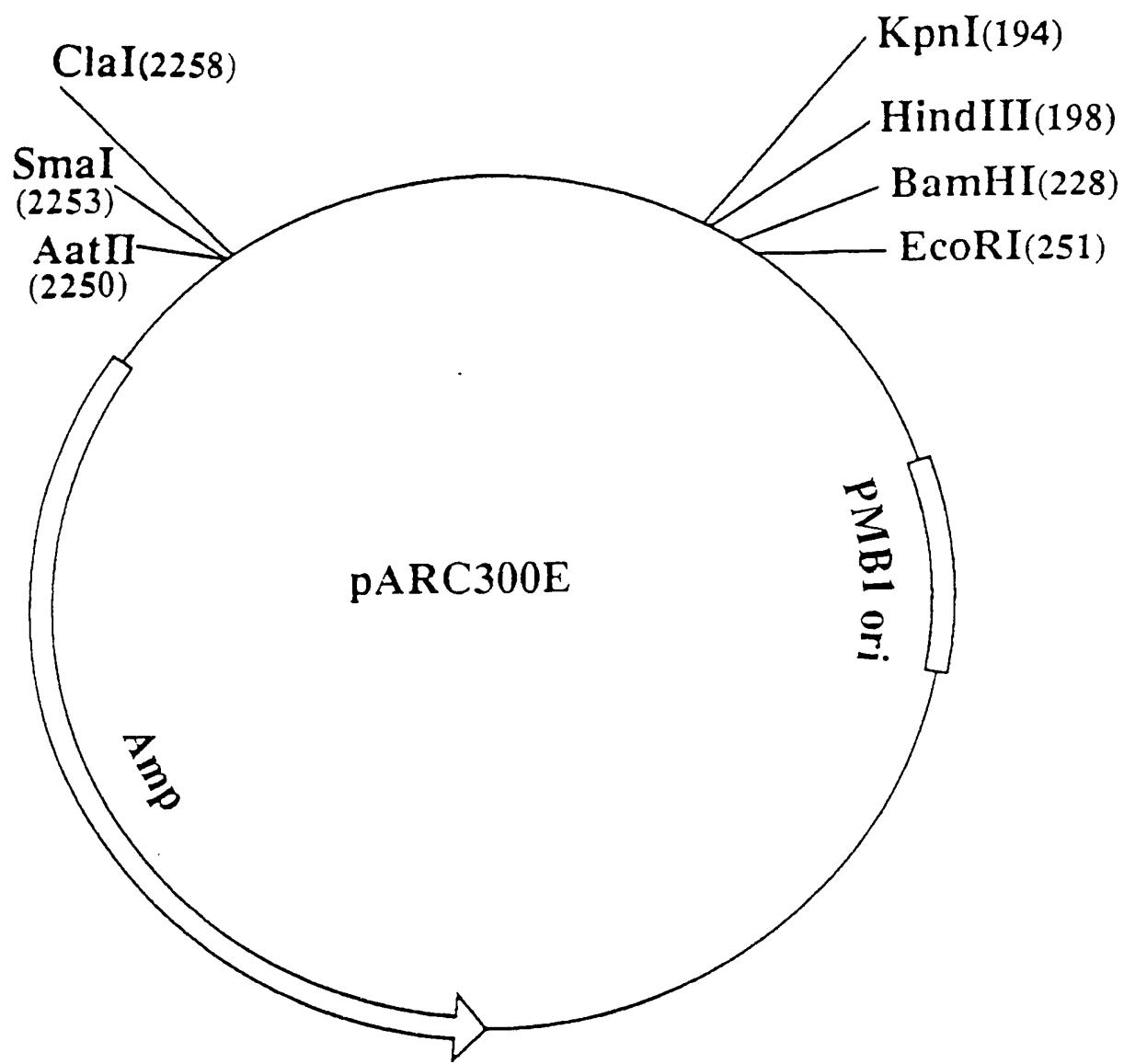TATTCGTCAGCTGGAAGGCCCAGCCCCCAGATCGCCAGGCGCCGATCCAGC

FIGURE 20-3

```
                                                                    60
Nco I
TGAGTCGGCAGGAGCGGCGCACCGGCCATGGTACTAAAATAGTTTAATCGTCATCTTGACCGTT
                        MetValLeuAsnSerLeuIleValIleLeuThrVal

                                                                   120
ATTGGCGATGGGAGGGCATCGCCGCGGTTTACCCCACCGCTACATTATGCCGGCTGGGGATGG
IleAlaMetGluGlyIleAlaIleTyrIleMetHisGlyTrpGlyTrp

                                                                   180
CGCTGGCATGAGCCACCACCATACCCCGCGCAAGGGCGTATTTGAGCTAAACGATCTCTTT
ArgTrpHisGluProHisHisThrProArgLysGlyValPheGluLeuAsnAspLeuPhe

                                                                   240
GCGGTGGTGTTGCCGGGTGGCTATCGCGCTGATTGCCGTGGGCACGGGCGGGGTTGG
AlaValValPheAlaGlyValAlaAlaIleAlaLeuIleAlaValGlyThrAlaGlyValTrp

                                                                   300
CCCCTGCAGTGGATTGGCGGCATGACGGTCTATGGCCGTCTCTACTTCCTGGTGCAC
ProLeuGlnTrpIleGlyCysGlyMetThrValTyrGlyLeuLeuTyrPheLeuValHis
                            Bam HI

                                                                   360
GACGGTCTGGTGCATCAGGCGCCCTTCCACTGGATCCCGCGCGGGGCTACCTGAAG
AspGlyLeuValHisGlnArgTrpProPheHisTrpIleProArgArgGlyTyrLeuLys

                                                                   420
CGGCTCTACGTCCCCACCGCCCTGCACCACGCGGCGCCGGAGGGCTGCCTCC
ArgLeuTyrValAlaHisHisAlaAlaValAlaArgGlyArgGluGlyCysValSer
```
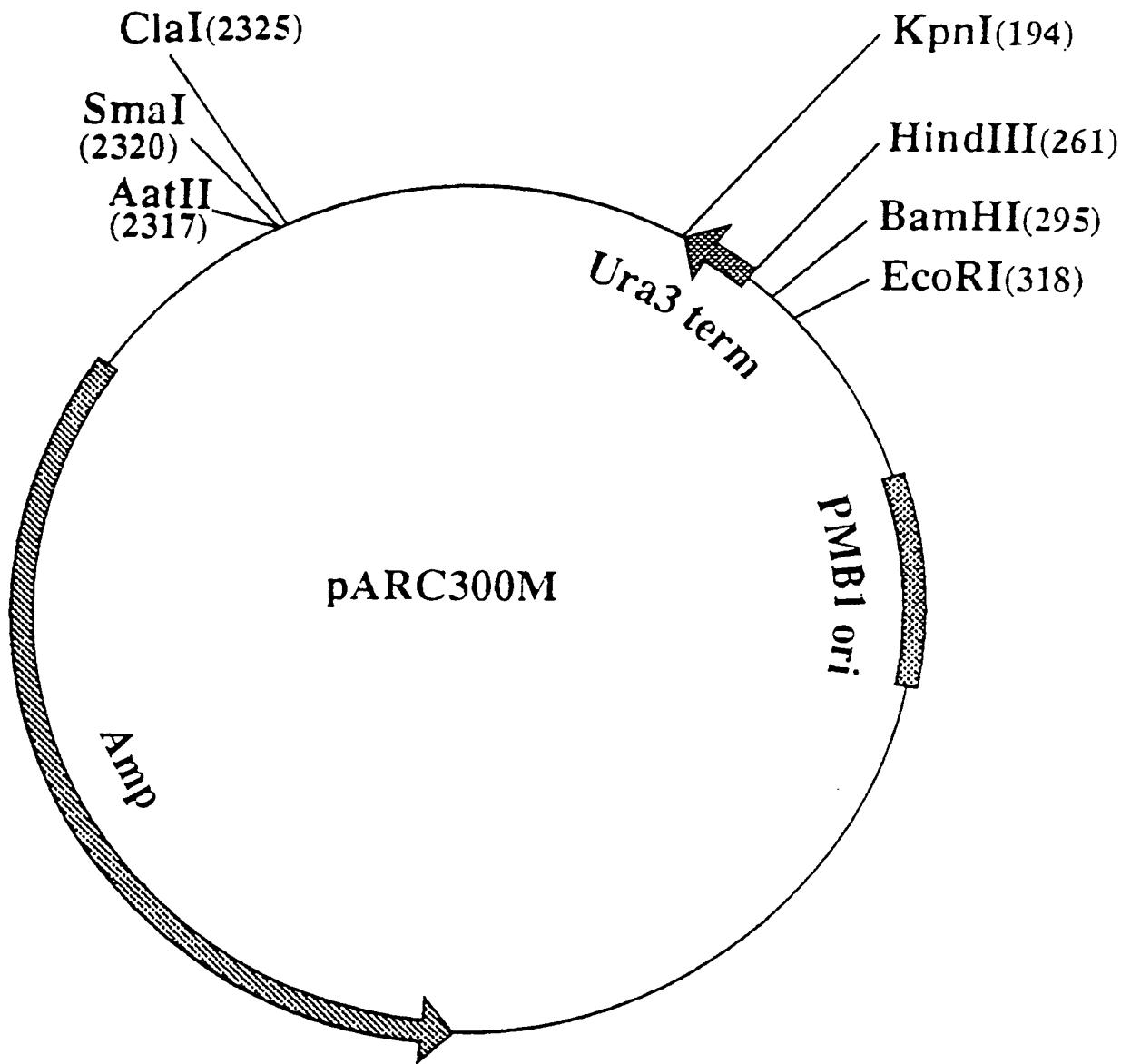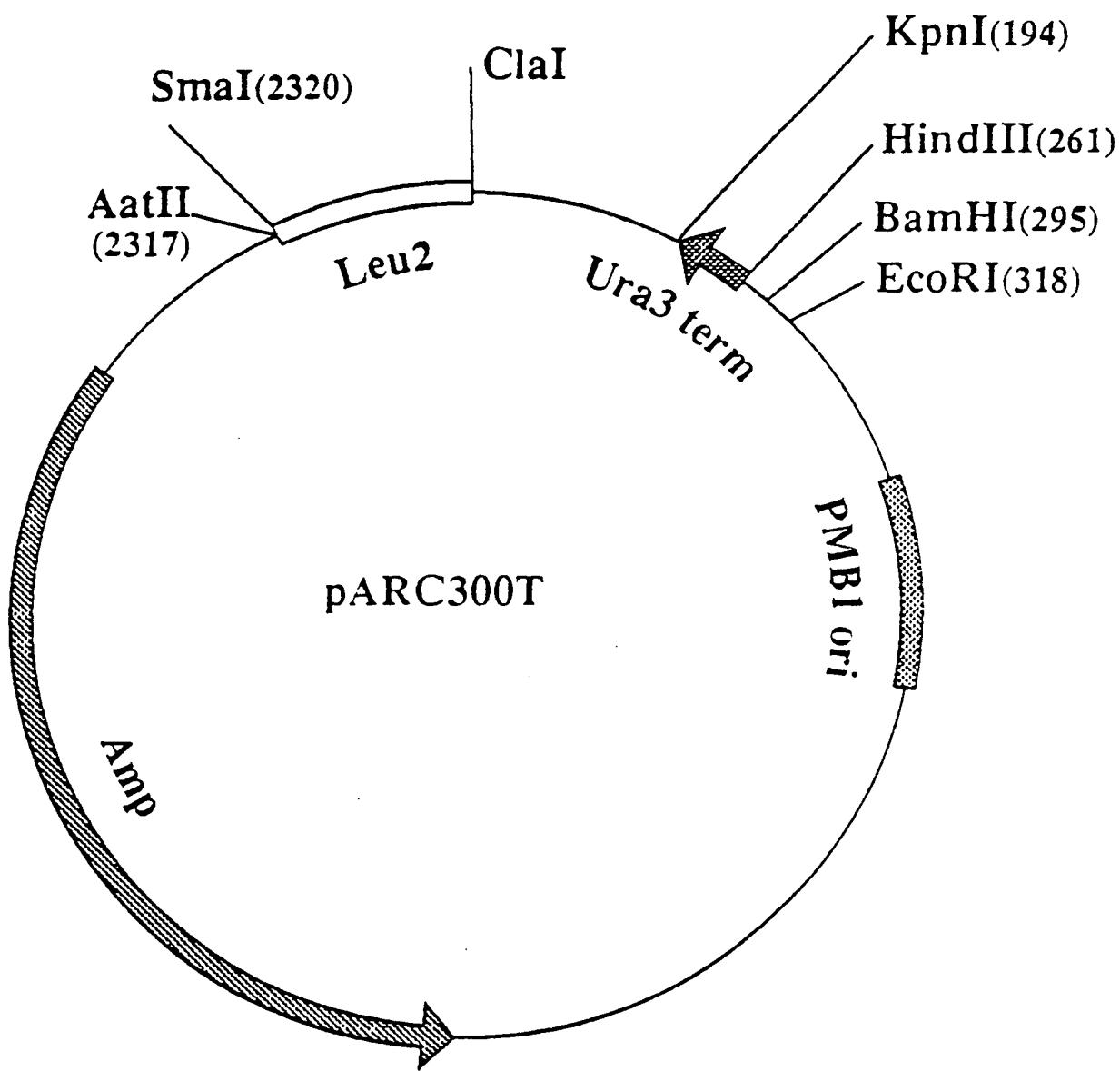
FIGURE 21-1

TTCGGTTTTATTTACGCCCGCCGACCTACAGGGCGATCCTGCCGTGAACGTCAT   480
PheGlyPheIleTyrAlaArgLysProAlaAspLeuGlnAlaIleLeuArgGluArgHis

GGCCGCCCGCCTAAACGGGACGCTGCCAAAGACCGGCGGCGTCACCCTCGTCG   540
GlyArgProProLysArgAspAlaAlaLysAspArgProAspAlaAlaSerProSerSer

TCTTCGCCCGAATAAACCTGCCCGGTGCCCGGCCATCAGCATGGCAATTTTTCACCTTTGC   600
SerSerProGlu

TGGTGTGCTGGGGCGATCCCCAGGCGCTCCCGGCTTTTACCTTAATACCGATCT   660

CCCGGTAGACGCTGGCCACGCGGGATCGCCCAGCGCCGGCAGATCGTGTA   720

GCCCGGCCCTGGAGGAGATGTAGTACGGCTCTGCGGCATCAATAAGCCTCCGCCACCGCG   780

CCAGCGCGGGGCGATTCTCCCGGCGGCCATAGTTCTCCCGGGGCCAGCCCGGCATCCTGCA   840

FIGURE 21-2

```
                                                    Sma I

GCCACTCGGGGGCAGATAGCAGCGGTCAATAGCCCGCATCGTCAATAATATCCCGGGCCA    900

TATTCGTCAGCTGGAAGGCCCAGCCCCCAGATCGCCAGGGCGCGATCCCAGC
```

FIGURE 21-3

FIGURE 22

41

FIGURE 23

FIGURE 24

43

```
ATGAGCCATTTGCCATTGTGGCACCCGCCCTCTACAGTCATGCGGTGGCGGTGCATGCC          60
METSerHisPheAlaIleValAlaProProLeuTyrSerHisAlaValAlaLeuHisAla

CTGGGCGCTGCAGATGGCCCAACGCGGCCACCGGGTGACCTTTCTCACCGGCAACGTCGCC         120
LeuAlaLeuGluGluMETAlaAlaLeuAlaLeuGlnThrGlyHisArgValThrPheLeuThrGlyAsnValAla

TCGCTGGCAGAGCAGGAAACGGAGCGGGTGGCGTTCTATCCCACTTCCCCGCCAGCGTGCAA         180
SerLeuAlaGluGlnGluThrGluArgValAlaPheTyrProLeuProAlaSerValGln

CAGGCCCCAGCGGCCAACGTCCAGCAGCAGAGTAACGGCAACCTGCTGGGCTGATTGCGGCC         240
GlnAlaGlnArgAsnValGlnGlnSerAsnGlyAsnLeuLeuArgLeuIleAlaAla

ATGTCATCCCTGACCGATGTGCTCTGCGAGCAGTTGCCCTATTCTACAGCGGTGGCG         300
METSerSerLeuThrAspValLeuCysGlnGlnLeuProAlaIleLeuGlnArgLeuAla

GTGGACGCGGCTGATTGTGTCGATGAGAGATGGAGCCCCGCCGGAAGCCTGGTCGCCGAGGCGCTG         360
ValAspAlaLeuIleLeuValAlaAspGluMETGluProAlaGlySerLeuValAlaAlaGluAlaLeu

GGACTACCATTTATCTCTATTGCCTGCGCGGTCAACCGGCAGCTGCCGCTGCCG         440
GlyLeuProPheIleSerIleAlaCysAlaLeuProValAsnArgGluLeuProLeuPro
```

FIGURE 25-1

500
GTGATGCCGTTCACTACGCCGAGGATAAGAGAGCCCGTGCCGTTTCAGGTCAGCGAA
ValMETProPheHisTyrAlaGluAspLysArgAlaArgAlaArgPheGlnValSerGlu

560
CGGATCTACGACGGCTGATGTACCCGGACGGGCCAGACGACCATCCTGCCGCCAGCGGC
ArgIleTyrAspAlaLeuMETTyrProHisGlyGlnThrIleLeuArgHisAlaGlnArg

620
TTTGGTTGCCGGAGCGGCCAGGCCGTCTCGACGAGTGTCTCTCGCCCCCTGGGCCAGATTAGC
PheGlyLeuProGluArgArgArgLeuAspGluCysLeuSerProLeuAlaGlnIleIleSer

680
CAGTCCGTTCCCGGCCTCGACTTCCCACGCCCGGCTGCCGAACTGTTTTACCTTACCGTG
GlnSerValProAlaLeuAspPheProArgAlaLeuProAsnCysPheThrTyrVal

740
GGAGCACTGCCTATCAGCCCCGCCGCTAGAACGCTCGCCAGCACGCCGCGG
GlyAlaLeuArgTyrGlnProProGlnValGluArgSerProArgThrProArg

800
ATCTTTGCCCTGCCTCCAGGGCCACCCTCCAGGGCGTCTACGCCGTCGTTTCAGAAGATCGCC
IlePheAlaSerLeuLeuGlyThrLeuGlnGlyHisGlyArgLeuArgLeuPheGlnLysIleAla

860
CGCGCCCTGTGCCAGCGTGGGCGCGGAGGTGACCATTGCCCACTGCGATGGCCGACGCCC
ArgAlaCysAlaSerValGlyAlaGluValThrIleAlaHisCysAspGlyLeuThrPro

FIGURE 25-2

GCCCAGGCCGACTCGCTCTACTGCGGCGGCGACGGAGGTGGTCAGCTTTGTCGACCAGCCG   940
AlaGlnAlaAspSerLeuTyrCysGlyAlaThrGLuValValSerPheValAspGlnPro

CGCTACGTTGCCGAGGCTAATCTGGTGATCACCCACGGGTCTCAATACCGTACTGGAT   1000
ArgTyrValAlaGluAlaAsnLeuValIleThrHisGlyLeuAsnThrValLeuAsp

GCGCTGGCCGCGACGCCGGTGCTGGCCGTGCCACTCTCTTTCGACCAGCCCGCCGTG   1060
AlaLeuAlaAlaAlaThrProValLeuAlaValProLeuSerPheAspGlnProAlaVal

GCTGCCCGGCTGGTCTATAACGGGCTGGGTCGCCCGGTATCGCGCTTGCCAGACAGGAG   1140
AlaAlaArgLeuValTyrAsnGlyLeuGlyArgArgValSerArgPheAlaAlaArgGlnGln

ACGCTGGGCGGATGAGATTGCCCAACTGCTGGGGGATGAGACGCTGCATGAGGCGTGTGGGCG   1200
ThrLeuAlaAspGluIleAlaGlnLeuLeuLeuGlyAspGluThrLeuHisGlnArgValAla

ACGGCCCAGCAGCTTAACGACGCCGGGGGCCACGCCCCGTTGCGGCGACCCTGATTGA   1260
ThrAlaArgGlnGlnLeuAsnAspAlaGlyGlyThrProArgCysGlyAspProAsp

FIGURE 25-3

I. CLASSIFICATI N OF SUB__CT MATTER (if several classification symbols apply, indicate all) 6
According to International Patent Classification (IPC) or to both National Classification and IPC
IPC(5): C07H 15/12; C12P 21/00, 23/00, 5/00; C12N 15/52,15/70,15/74,15/81
US CL: 536/27; 435/67,69.1,166,172.3,320

II. FIELDS SEARCHED

Minimum Documentation Searched 7	
Classification System	Classification Symbols
U.S.	435/67,69.1, 166,172.3,320; 536/27; 935/35,60,64,67,69,72,73

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched 8
U.S. PTO Automated Patent System; DIALOG files: BIOTECH, PATENTS; Gen BANK/UEMB L and PIR/SWISS PROT See Attachment for search terms

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
Y	Plants Tody; Volume 1, Number 1, Issues January-February 1988, Bryant, "Putting genes into plants", pages 23-28. See page 26.	5,12,16,23, 25,28,38,42 52,56,65,69 and 74-76
Y	Weed Science, Volume 26, Issued 1978, Bartels et al, "Inhibition of Carotenoid Synthesis by fluridone and norflurazon", pages 198-203. See entire document.	74-76
Y	Plant Physiology (Bethesda, USA), Volume 88, Number 2, Issued 1988, Sagar et al," Light effects on several chloroplast components in norflurazon-treated pea seedlings", pages 340-347. See entire document.	74-76

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
29 May 1991	07 AUG 1991
International Searching Authority	Signature of Authorized Officer
ISA/US	P.Rhodes

Form PCT/ISA/210 (second sheet) (Rev.11-87)

Attachment to Form PCT/ISA/210, Part II.

II.   FIELDS SEARCHED SEARCH TERMS:

norflurazon resistance; carotenoid or geranylgeranyl or phytoene or lycopene or carotene or zeaxanthin; erwinia; Rhodobacter; sequence or clone.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECO[ ] SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X Y	Journal of Biological Chemistry; Volume 264, Number 22; Issued 05 August 1989; Bartley et al; "Carotenoid biosynthesis in photosynthetic bacteria genetic characterization of the Rhodobacter capsulatus CrtI protein"; pages 13109-13113. (see entire document)	19-22, 24 33, 74-76
Y,P	Journal of Biological Chemistry; Volume 265, Number 14; Issued 15 May 1990; Armstrong et al; "Genetic and biochemical characterization of carotenoid biosynthesis mutants of Rhodobacter capsulatus"; pages 8329-8338. (see entire document)	1-73
Y,P	Journal of Biological Chemistry; Volume 265, Number 26; Issued 15 September 1990; Bartley et al; "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora crassa are structurally and functionally conserved and contain domains homologous to flavoprotein disulfide oxidoreductases"; pages 16020-16024. (see entire document)	1-73
Y	European Journal of Biochemistry; Volume 184, Number 2; Issued September 1989; Schmidt et al; "Immunological detection of phytoene desaturase in algae and higher plants using an antiserum raised against a bacterial fusion-gene construct"; pages 375-378. (see entire document)	1-76
Y	Journal of General Microbiology; Volume 130; Issued 1984; Thiry; "Plasmids of the epiphytic bacterium Erwinia uredovora"; pages 1623-1631. (see entire document)	1-76
X, P Y	Gene; Volume 91, Number 1; Issued 02 July 1990; Schmidt et al; "Cloning and nucleotide sequence of the crtI gene encoding phytoene dehydrogenase from the cyanobacterium Aphanocapsa PCC6714"; pages 113-117. (see entire document)	19-22, 24 18,25-33 44-46,57-59 70-72,74-76
Y,P	Proceedings of the National Academy of Sciences USA; Volume 87; Issued December 1990; Armstrong et al; "Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes"; pages 9975-9979. (see entire document)	1-73

III. DOCUMENTS CONSIDERED 3E RELEVANT (CONTINUED FROM THE SECOND S ET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X,P	FEMS Microbiology Letters; Volume 78; Issued 01 March 1991; Schnurr et al; "Mapping of a carotenogenic gene cluster from Erwinia herbicola and functional indentification of six genes"; pages 157-162. (see entire document)	1-4,8-11,19-22, 34-37,48-51, and 61-64
Y	Journal of Cellular Biochemistry; Volume 12C; Issued 1988; Bennetzen et al; "Structure and protective role of the carotenoid synthesis gene(s) of Erwinia stewartii"; page 246. (see abstract Y101)	1-76
Y	Trends in Genetics; Volume 4, Number 8; Issued August 1988; Botterman et al; "Engineering herbicide resistance in plants"; pages 219-222. (see entire document)	74-76
Y	Bio/Technology; Volume 6; Issued August 1988; Hinchee et al; "Production of transgenic soybean plants using Agrobacterium-mediated DNA transfer"; pages 915-922. (see entire document)	74-76
X / Y	Molecular and Cellular Biology; Volume 9, Number 3; Issued March 1989; Nelson et al; "Molecular cloning of a Neurospora crassa carotenoid biosynthetic gene (Albino-3) regulated by blue light and the products of the white collar genes"; pages 1271-1276. (see entire document)	1,3-4,6 / 2,5,7,14-17, 29-32,44-46, 57-59,70-72

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X Y	Current Microbiology, Volume 15, Issued 1987, Pemberton et al, "Expression of Rhodopseudomonas sphaeroides carotenoid photopigment genes in phylogenetically related nonphotosynthetic bacteria", pages 67-71. See entire document.	1,3-4,6,8,10 11,13-15,19 21-22,24,26-27,29,34,36-37,39-41,43-44,48,50-51, 53-55,57,61, 63-64,66-68, 70
		2,5,7,9,12,16 18,20,23,25, 28,30-33,35, 38,42,45-47, 49,52,56,58-60,62,65,69 71-73
Y	Phytopathology, Volume 79, Number 2, Issued February 1989, Daub et al, "The role of carotenoids, in resistance of fungi to cercosporin", pages 180-185, See entire document.	74-76
Y	Journal of Phycology, Volume 23, Number 1, Issued 1987, Ben-Amotz et al, "Massive accumulation of phytoene induced by norflurazon in <u>Dunaliella-bardawil</u> (chlorophyceae) prevents recovery from photoinhibition", pages 176-181. see entire document.	74-76

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND EET)		
Category *	Citation of Document... with indication, where appropriate, of the relevant passages	Relevant to Claim No
$\underline{X}$ Y	Journal of Bacteriology; Volume 168, Number 2; Issued November 1986; Perry et al; "Cloning and Regulation of Erwinia herbicola Pigment Genes"; pages 607-612. (see entire document)	1-4,6-11,13-15,17-22, 24,26-27,29-30,33-37, 39-41,43-45,47-51, 53-55,57-58,60,64, 66-68,70-71, and 73 5,12,16,23,25,28 31-32,38,42,46,52, 54-56,59,65,69,72,74-76
Y	Journal of Bacteriology; Volume 170, Number 10; Issued October 1988; Tuveson et al; "Role of Cloned Carotenoid Genes Expressed in Escherichia coli in Protecting against Inactivation by Near-UV Light and Specific Phototoxic Molecules"; pages 4675-4680. (see entire document)	1-76
A	Journal of Bacteriology; Volume 171, Number 9; Issued September 1989; Tichy et al; "Genes Downstream from pucB and pucA are Essential for Formation of the B800-850 Complex of Rhodobacter capsulatus"; pages 4914-4922. (see entire document)	1-76
X,P	EP A 0,393,690 (MISAWA ET AL) 24 October 1990. (see entire document)	1-76
$\underline{X}$,P Y	Journal of Bacteriology; Volume 172, Number 12; Issued December 1990; Misawa et al; "Elucidation of the Erwinia uredovora Carotenoid biosynthetic pathway by functional analysis of gene products expressed in Escherichia coli"; pages 6704-6712. (see entire document)	1-4,6-11,13-15,17-22, 24,26-27,29-30,33-37, 39-41,43-45,47-51, 53-55,57-58,60,64, 66-68,70-71, and 73 5,12,16,23,25,28 31-32,38,42,46,52, 54-56,59,65,69,72,74-76
Y	Molecular and General Genetics; Volume 213, Number 1; Issued July 1988; Giuliano et al; "A genetic-physical map of the Rhodobacter capsulatus carotenoid biosynthesis gene cluster"; pages 78-83. (see entire document)	1-76
$\underline{X}$ Y	Molecular and General Genetics; Volume 216, Number 2/3; Issued April 1989; Armstrong et al; "Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of Rhodobacter capsulatus"; pages 254-268. (see entire document)	1,3-6,8,10-19, 21-34,36-48, 50-61, and 63-73 2,7,9,20,35,49,62
A	Molecular and General Genetics; Volume 218, Number 1; Issued July 1989; Young et al; "Genetic evidence for superoperonal organization of genes for photosynthetic pigments and pigment-binding proteins in Rhodobacter capsulatus"; pages 1-12. (see entire document)	1-76

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	Zeitschrift fur Naturforschung, Volume 45c, Issued March 1990, Chamovitz et al, "Cloning a gene coding for norflurazon resistance in cyanobacteria" pages 482-486. See entire document.	74-76

V.☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE¹

This International search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1.☐ Claim numbers ..... because they relate to subject matter ¹² not required to be searched by this Authority, namely:

2.☐ Claim numbers ...., because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out ¹³, specifically:

3.☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI.☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING²

This International Searching Authority found multiple inventions in this international application as follows:

See Attachment

1.☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application. Telephone Practice.

2.☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3.☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4.☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (supplemental sheet (2) (Rev. 11-87)

VI. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

The claims present a plurality of mutually exclusive independent inventions as follows:

I. Claims 1-7 and 17, drawn to a structural gene for geranylgeranyl pyrophosphate synthase, and vectors containing same and a first method of use to make geranylgeranyl pyrophosphate synthase protein, classified in Classes 435 and 536, subclasses 69.1 and 320 and subclass 27, respectively, for example.

II. Claims 8-12 and 18, drawn to a structural gene for phytoene synthase and vectors containing same, classified in Class 435 and 536, subclasses 320 and 27, respectively, for example.

III. Claim 13, drawn to a method of use for the product of Group II to make phytoene synthase protein, classified in Class 435, subclass 69.1, for example.

IV. Claims 14-16, drawn to a method of making the compound phytoene, classified in Class 435, subclass 166, for example.

V. Claims 19-23 and 33, drawn to a structural gene for phytoene dehydrogenase-4H and vectors containing same, classified in Class 435 and 536, subclasses 320 and 27, respectively, for example.

VI. Claims 24-25, drawn to a first method of use for the product of Group V to make phytoene dehydrogenase-4H protein, classified in Class 435, subclass 69.1, for example.

VII. Claims 26-28, drawn to a first method of making lycopene, classified in Class 435, subclass 166, for example.

VIII. Claim 29-32, drawn to a second method of making lycopene, classified in Class 435, subclass 166, for example.

IX. Claims 34-38 and 47, drawn to a structural gene for lycopene cyclase and vectors containing same, classified in Class 435 and 536, subclasses 320 and 27, respectively, for example.

X. Claim 39, drawn to a method of use for the product of Group IX to make lycopene cyclase protein, classified in Class 435, subclass 69.1, for example.

XI. Claim 40-43, drawn to a first method of making beta-carotene, classified in Class 435, subclass 67, for example.

XII. Claim 44-46, drawn to a second method of making beta-carotene, classified in Class 435, subclass 67, for example.

XIII. Claims 48-52 and 60, drawn to a structural gene for beta-carotene hydroxylase and vectors containing same, classified in Class 435 and 536, subclasses 320 and 27, respectively, for example.

VI.   OBSERVATIONS   WHERE   UNITY   OF   INVENTION   IS   LACKING
(continuation)

    XIV.   Claim 53, drawn to a method of use for the product of Group XIII to make beta-carotene hydroxylase protein, classified in Class 435, subclass 69.1, for example.

    XV.   Claims 54-56, drawn to a first method of making zeaxanthin, classified in Class 435, subclass 67, for example.

    XVI.   Claim 57-59, drawn to a second method of making zeaxanthin, classified in Class 435, subclass 67, for example.

    XVII.   Claims 61-65 and 73, drawn to a structural gene for zeaxanthin glycosylase and vectors containing same, classified in Class 435 and 536, subclasses 320 and 27, respectively, for example.

    XVIII.   Claim 66, drawn to a method of use for the product of Group XVII to make zeaxanthin glycosylase protein, classified in Class 435, subclass 69.1, for example.

    XIX.   Claims 67-69, drawn to a first method of making zeaxanthin diglucoside, classified in Class 435, subclass 67, for example.

    XX.   Claim 70-72, drawn to a second method of making zeaxanthin diglucoside, classified in Class 435, subclass 67, for example.

    XXI.   Claims 74-76, drawn to a second method of use for the product of Group V to make plants resistant to the herbicide norflurazon, classified in Class 435, subclass 172.3, for example.